2018

Analysis of Remote Tripping Command Injection Attacks in Industrial Control Systems Through Statistical and Machine Learning Methods

Charles Timm
University of Central Florida

ANALYSIS OF REMOTE TRIPPING COMMAND INJECTION ATTACKS IN INDUSTRIAL CONTROL SYSTEMS THROUGH STATISTICAL AND MACHINE LEARNING METHODS

by

CHARLES R. TIMM
B.S. United States Military Academy, 2008
M.S. University of Central Florida, 2018

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Modeling and Simulation
in the Department of Modeling and Simulation
in the College of Graduate Studies
at the University of Central Florida
Orlando, Florida

Summer Term
2018

Major Professor: Bruce Caulkins

ABSTRACT

In the past decade, cyber operations have been increasingly utilized to further policy goals of state-sponsored actors to shift the balance of politics and power on a global scale. One of the ways this has been evidenced is through the exploitation of electric grids via cyber means. A remote tripping command injection attack is one of the types of attacks that could have devastating effects on the North American power grid. To better understand these attacks and create detection axioms to both quickly identify and mitigate the effects of a remote tripping command injection attack, a dataset comprised of 128 variables (primarily synchrophasor measurements) was analyzed via statistical methods and machine learning algorithms in RStudio and WEKA software respectively. While statistical methods were not successful due to the non-linearity and complexity of the dataset, machine learning algorithms surpassed accuracy metrics established in previous research given a simplified dataset of the specified attack and normal operational data. This research allows future cybersecurity researchers to better understand remote tripping command injection attacks in comparison to normal operational conditions. Further, an incorporation of the analysis has the potential to increase detection and thus mitigate risk to the North American power grid in future work.

ACKNOWLEDGMENTS

First and foremost, I want to thank my wife Amelie for the support, patience, and putting up with the many nights of not enjoying the beautiful Florida weather due to this academic endeavor. I'd also like to thank my thesis advisor, Dr. Bruce Caulkins. Your guidance and encouragement throughout this process has been absolutely crucial for completion, and I greatly appreciate the countless meetings and emails exchanged over the course of these past two years. I'd also like to thank the members of my thesis advisory committee, Dr. Wiegand and Dr. Lathrop. Dr. Wiegand, your help with statistical methods and machine learning algorithms for the last six months has been integral in the construction of this thesis. Thanks for your patience, and ability to explain complex subject matter in a way that is both meaningful and easy to understand and integrate into this work. Dr. Lathrop, thanks for providing insight about my methodology and asking the questions which ultimately led me in the right direction regarding thinking about the research questions and formulating a solution. I'd also like to acknowledge Dr. Owen Wilson and Dr. Tommy Morris, subject matter experts (SMEs) in Industrial Control Systems (ICS) and my sounding board regarding all domain related questions; without your knowledge and the construction of this dataset, this thesis would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AIC – Akaike Information Criterion

ARFF – Attribute Relation File Format

CMF – Cyber Mission Force

CPS – Cyber Physical System

CSV – Common Separated Values

DHS – Department of Homeland Security

DoD – Department of Defense

HMI – Human Machine Interface

FP – False Positive

FN – False Negative

GINA – Global Information Network Architecture

ICS – Industrial Control System

ICS-CERT – Industrial Control Systems Cyber Emergency Response Team

IDS – Intrusion Detection System

IED – Intelligent Electronic Device

IT – Information Technology

ML – Machine Learning

NERC – North American Electric Reliability Corporation

NERC-CIP – North American Electric Reliability Corporation Critical Infrastructure Protection

NIST – National Institute of Standards and Technology

OT – Operational Technology

PCA – Principal Component Analysis

PDC – Phasor Data Concentrator

PLC – Programmable Logic Controller

PMU – Phasor Measurement Unit

RQ – Research Question

RTA – Russian Threat Actors

RTU – Remote Terminal Unit

NNGe – Non Nested Generalized Exemplars (nearest neighbor like ML algorithm)

SCADA – Supervisory Control and Data Acquisition

SME – Subject Matter Expert

SLR – Stepwise Logistic Regression

TT – Time Taken

TTP – Tactics, Techniques, and Procedures

US – United States

UTC – Universal Time Coordinated

VRDM – Vector Relational Data Modeling

WAMS – Wide Area Monitoring System (WAMS)

WAN – Wide Area Network

WEKA – Waikato Environment for Knowledge Analysis

CHAPTER 1: PURPOSE/RESEARCH MOTIVATION

This chapter provides an overview of the concepts that will be discussed in the body of this research. The major themes, techniques, and technologies will be presented in a macro sense to illustrate the motivation behind the research, as well as a synopsis and direction of the remaining sections. The majority of this introductory section is devoted to providing the motivation and inspiration behind this work, which is best demonstrated through an overview of the current state of geopolitics and recent developments amongst the inextricably linked domains of the energy sector and cyber operations. The ultimate objective of this chapter is to provide a base level of background information to facilitate understanding of the research questions and to demonstrate there is a credible threat looming with the potential to strike the North American electric grid. The state of current affairs suggests this field of research is meaningful and necessary to bolster the security of critical infrastructure in the United States (US).

1.1 Introduction – The State of Cyber Operations in Geopolitics and Relation To Industrial Control Systems/Cyber Physical Systems

In the past two years, the US government, military, and several private cybersecurity organizations have published documents detailing the rising risk of

malicious cyber operations. In 2016, the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT, an organization of the US government which falls under the Department of Homeland Security (DHS)) responded to over 290 incidents, of which 59 were in the energy sector, 62 were in the communications sector, and 63 in the critical manufacturing sector (ICS-CERT, 2016). Of note, spear phishing was present in 26% of the incidents making it the leading access vector, and the first known cyberattack to result in a physical impact to a power grid was observed (ICS-CERT, 2016). 2017 was by all accounts a record breaking year in that there were 159,700 reported cyber incidents, an 18.2% increase in reported breach incidents, a $5 billion financial impact from ransomware, and a 90% rise in business targeted ransomware (Online Trust Alliance, 2018).

With this notable increase of cyber-attacks in the civilian sector and the evolution of state actors weaponizing cyber operations to both disrupt adversaries on the battlefield and influence geopolitics, in 2017 the US Department of Defense (DoD) increased the capabilities of US Cyber Command to a unified combatant command under Title 10, which means the unit is legally capable of conducting offensive cyberspace combat operations (Department of Defense, 2017). The need to develop cyber capabilities has resulted in significant increases in funding and directives aimed at growing and educating US military cyber operations personnel, or what is now referred to as the Cyber Mission Force (CMF). The CMF's three primary missions include 1) defending and securing DoD networks and data, 2) supporting joint military commander objectives, and 3) defending U.S. critical infrastructure (Pomerleau, 2016).

The total DoD CMF is comprised of 133 teams and 6200 personnel (task organization: 13 teams to defend the nation's infrastructure, 68 to defend DoD networks, 27 to provide support to combatant commanders, and 25 to provide analytic and planning support to the teams) (Pomerleau, 2017). As indicated above, there are significant resources and dedicated cyber defense teams being allocated to critical infrastructure in the US. This is mirrored in Strategic Goal II of the DoD Cyber Policy which states, "Be prepared to defend the U.S. homeland and U.S. vital interests from disruptive or destructive cyberattacks of significant consequence" (Department of Defense, 2015). The thirteen teams tasked with defense of the nation's infrastructure face a multitude of threats, of which many are focused on the denial of essential services to the populace.

In addition to cyber operations being a powerful overt and conventional weapon during open conflict between nation states, disruption of essential services such as water, electricity, and natural gas can be wielded by adversarial nations as a tool to erode trust in a government's legitimacy and ability to provide for its citizens. The cyber dimension adds layers of complexity due to difficulties in proving decisive attribution to a particular actor. In a way that mirrors what many pundits feel was malicious intent exhibited through likely nation state interference in the 2016 US Presidential elections, a possible ultimate goal of a malicious actor is to decrease citizens' faith in the system itself, rather than the immediate effect of denying services to the population. However, the denial of essential services such as electricity would also likely have severe economic implications due to reliance on electrically powered devices that form the

foundation of local business, and undoubtedly greatly affect the lives of citizens reliant on personal devices.

There is evidence to suggest that Russia has conducted a proof of concept of this type of attack by depriving essential services in the Ukraine via a cyber vector in 2015 and 2016 that denied power to close to half a million Ukrainians (Greenberg, 2017). These events could potentially be a proof of concept or test bed for future engagements with nations that have greater cybersecurity capabilities/risk mitigation (Greenberg, 2017). Denial or disruption of electricity could be used in concert with combat operations against a conventional opponent but could potentially be more effective against the civilian population to sow seeds of distrust and doubt about the efficacy of the government in responsibility of the affected region. This could be seen as a larger campaign to erode citizens faith in the targeted government. In any case, there is an immense need to classify attacks against the power grid quickly and with a high degree of accuracy to identify malicious activity early and mitigate damage/disruption of critical infrastructure and essential services.

The delivery of these essential services is made possible by Industrial Control Systems (ICS), which is a general term applied to control systems found in industrial sectors/critical infrastructure (NIST, 2015). ICS are comprised of myriad interoperating Information Technology (IT) and Operational Technology (OT) components that act together to achieve an industrial objective and facilitate distribution of services to the population (NIST, 2015; Murphy, 2017). This interoperation has multiple potential issues; first, there are systems that are interacting that were not initially designed to do

so; second, many of these interactions with legacy systems have outdated or unpatched security protocols; and third, is that there are generally conflicting interests between IT and OT processes/personnel, in which the OT side (usually comprised of engineers and management) are concerned with delivering service to a customer and the IT side (primarily comprised of cybersecurity or IT personnel) being concerned with security (and keeping systems and software updated to facilitate that security) (Murphy, 2017). There is a natural divide as reliability and security can often collide and decisions must be made that could hinder either side. There are of course ramifications if either is neglected indefinitely, or if management consistently prioritizes one over the other (Murphy, 2017). This dichotomy of competing interests can lead to vulnerabilities in the system that can in turn be potentially exploited by sophisticated threat actors. While these threats have been present for many years, recent events on the global stage suggest such as the Ukrainian energy sector attacks that “potential” ICS vulnerabilities have been exploited by state sponsored actors via cyber operations.

In the past three months, there has been increasing evidence to suggest that a nation state has infiltrated the North American power grid and has the potential to execute malicious follow on operations. This was shown in recent testimony in which the Secretary of the Department of Energy Rick Perry testified at a congressional hearing that there has undoubtedly been Russian infiltration in the North American power grid (Grandoni, 2018). The United States Computer Emergency Readiness Team (US-CERT) echoed this congressional testimony in a publicly acknowledged (which is a new precedent) joint technical alert spearheaded by the Department of

Homeland Security (DHS) and the Federal Bureau of Investigation (Alert TA18-07A (Russian Government Cyber Activity Targeting Energy and Other Critical Infrastructure Sectors). The alert warned of an ongoing multi-stage intrusion campaign by Russian Threat Actors (RTA) and that the group has moved laterally throughout peripheral and target networks and are actively conducting network reconnaissance and information regarding North American ICS infrastructure (US-CERT, 2018).

The US-CERT report detailed that this campaign had been active since at least March of 2016, and that in addition to the energy sector, targeting also occurred in the nuclear, water, aviation, and other critical manufacturing sectors (US-CERT, 2018). The campaign was executed via the use of common tactics, techniques, and procedures (TTPs) such as spear-phishing, staging of malware, and credential gathering, and initial victims were peripheral organizations with less secure networks but with access to the intended targets networks, which were then subsequently targeted (US-CERT, 2018). Additionally, common ICS trade publication/informational websites were modified to include malicious content with the ultimate goal of gaining credentials by compromising the peripheral organizations and an end goal of compromising an ICS organizational network and thus ICS infrastructure (US-CERT, 2018). Once the intended target network was infiltrated, there were multiple observations of workstations and servers being accessed that contained data output from control systems within ICS facilities (US-CERT, 2018). Files including (but not limited to) the configuration of systems within a specific ICS environment and Human Machine Interfaces (HMI) were accessed by the threat actor, indicating a high level of sophistication and information gain from the

campaign including reconstructed screenshots of exploited HMI accessed by RTA (US-CERT, 2018). While obtaining these screen captures, in at least one instance RTA used a logical format and naming convention that indicated the machine description, machine location, and organization name (Symantec, 2017). Additionally, a string of "cntrl" was noted in some machine descriptions, potentially marking the machine as those the RTA has gained access to (Symantec, 2017).

Leadership from prominent cybersecurity research organizations in the private sector have concurred with these assessments by the US government, with Eric Chien, a security technology director at Symantec stating in reference to Russia, "We now have evidence they're sitting on the machines, connected to industrial control infrastructure, that allow them to effectively turn the power off or effect sabotage," (Perlroth and Sanger, 2018). A report from Symantec named the group of attacks the Dragonfly and Dragonfly 2.0 campaigns and asserts that RTA have been in operation since at least 2011, with possible attributions/involvement in notable energy sector attacks such as the 2015 and 2016 Ukraine power grid attacks mentioned above (Symantec, 2017). Although the extent and scope of the infiltration of the North American grid by RTA is unknown, there is substantial evidence to suggest that the group is focused on gaining access to operational systems within the energy sector which could in turn be used to disrupt essential services (Symantec, 2017).

Therefore, both government and private sector entities concur, based off the most recent evidence, that there exists a high likelihood that the systems which regulate and provide electricity on the North American Electric Grid have been compromised by

RTA. If ICS networks have already been compromised (which is assumed in the dataset utilized in this work), the US is already in a reactionary state. However, there is still a need to understand when an attack is occurring as quickly as possible to decrease the severity and reduce impacts to the population. The use of synchrophasor technology and measurements via sensors will be explored in later sections (as it comprises the majority of data in the dataset) and should be a key tenant in a hybrid Intrusion Detection System (IDS) comprised of both physical processes within ICS and the network/cyber component to quickly discern between attacks and normal operational conditions and thus mitigate damage to the greatest extent possible.

To aid in the classification of attacks against ICS, a specific type of attack that has the propensity to disrupt the power grid through the exploitation of components associated with a generator, referred to as a remote tripping command injection attack, will be analyzed and modeled in the remaining sections of this work. The purpose of this research is to utilize statistical and/or machine learning methods to develop detection axioms for this type of attack via the analysis of variables (and parameters of variable values) in an open source dataset, and ultimately contribute to the formulation of an Intrusion Detection System (IDS) to thwart attacks employed against ICS. After finding an appropriate open source dataset, email communication with one of the dataset authors (Dr. Tommy Morris), indicated there was a dearth of open source research focused purely on detection of remote tripping command injection attacks (to his knowledge and further confirmed via the search of numerous academic journal databases), and thus was identified as an area in which the author could contribute to

the existing body of knowledge (Morris, 2018). The utilization of simple methods was identified as a logical starting point to achieve baseline results and were applied prior to application of more sophisticated machine learning (ML) methods. This research is focused on identifying a specific type of attack against the power grid with the highest accuracy possible, which could then potentially be utilized in the construction of a hybrid Intrusion Detection Systems (IDS) utilizing synchrophasor measurements in ICS. The following research questions were thus developed to focus this thesis and facilitate solution formation.

1.2 Research Question 1 (RQ1)

Can the simple statistical methods of Logistic Regression, Stepwise Logistic Regression, or Principal Component Analysis classify remote tripping command injection attacks on electrical grids at 95% or greater accuracy?

1.3 Research Question 2 (RQ2)

Can machine learning methods classify remote tripping command injection attacks on electrical grids at 95% or greater accuracy?

1.4 Research Question 3 (RQ3)

If RQ1 or RQ2 is true, can detection axioms be derived from this data for implementation in an IDS in future work?

1.5 Research Question 4 (RQ4)

In future work can these axioms/results be implemented in a Global Information Network Architecture (GINA) based IDS?

1.6 Solution Formation

Through an analysis of the Aurora event to define and explain a remote tripping command injection attack and a synthesized dataset (created by Mississippi State University and Oak Ridge National Laboratory and modified by the author to incorporate only remote tripping command injection attacks and normal operational data points) to analyze remote tripping command injection attack and normal operational data, the author will develop detection axioms based off confirmed relationships in the data utilizing statistical/ML methods. It should be noted, that the original direction/goal of this thesis was to utilize GINA, a vector relational data modeling (VRDM) approach to

classify the data via multi-attribute scripts within the system to further analyze/classify remote tripping command injection attacks (and possibly compare/contrast the metrics of statistical/ML methods to determine which method yields better results). Key to this utilization of GINA was the development of detection axioms which could be described as algorithms/rules to determine whether a given instance in the dataset was an attack or normal operations. These detection axioms/rules would form the basis of a conceptual model for integration into the GINA framework. Although GINA implementation was not realized in the course of the research largely due to the complexity of the dataset and time taken to analyze it, there is evidence to suggest that GINA could be incorporated in future work. The timeline/methodology below in Figure 1 demonstrates the key task breakdown for the research and a general synopsis of each task as originally envisioned and briefed to thesis advisory committee on 22FEB18.

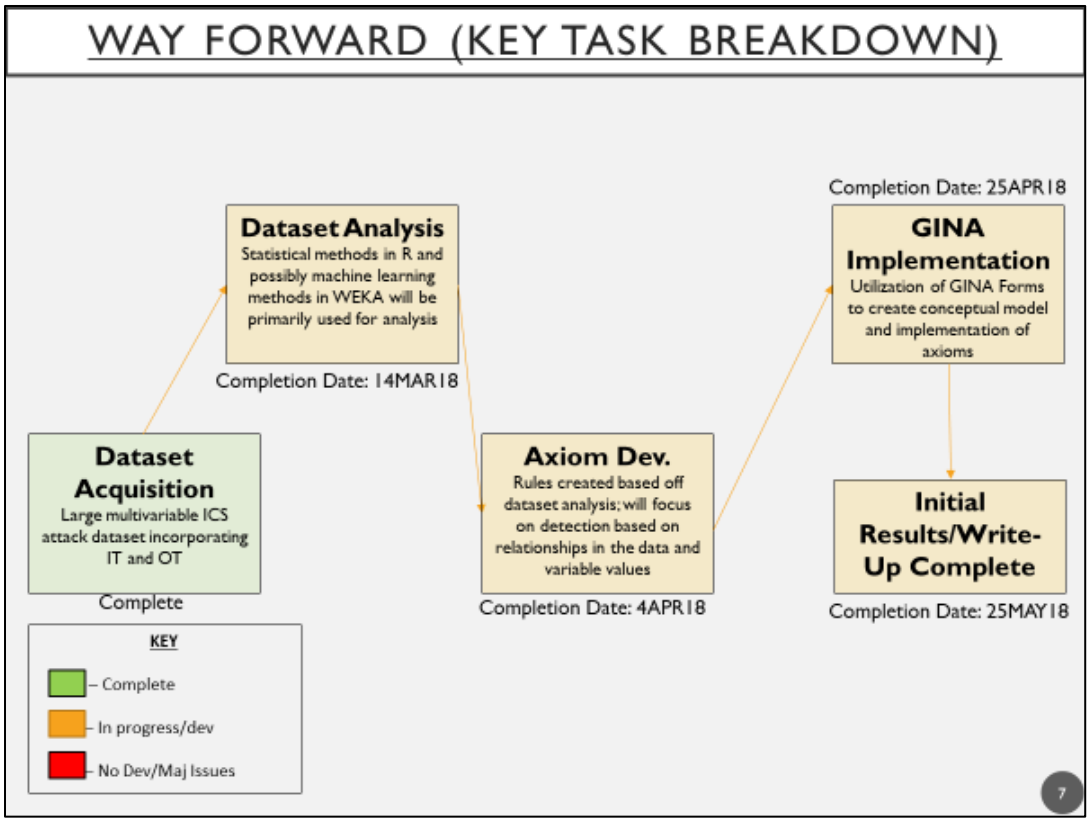


Figure 1 - Thesis Key Task Breakdown

1.7 Thesis Organization

The sections are organized as follows to facilitate understanding of this thesis: Section II consists of background information regarding ICS and Cyber Physical Systems (CPS), the Aurora Vulnerability and implications of remote tripping command injection attacks, and synchrophasor technology and its integration into IDS. Section III focuses on the synthesis and modification of the dataset, previous work associated with the dataset, statistical methodology/approaches utilized with results, and the ML

methodology/approaches utilized with results. Section IV summarizes the results and conclusions about utilization of more sophisticated methods to classify attacks given the dataset, issues/limitations of the research, and possible future work associated with this research.

CHAPTER 2: BACKGROUND INFORMATION

This section provides the reader with the basic concepts of electrical power generation and the various machinery/components that make it possible, the concept of Information Technology (IT) vs Operational Technology (OT) in an electrical grid context, synchrophasor technology, and a review the Aurora Event and remote tripping command injection attacks.

2.1 Electrical Power Generation, Transmission, and Distribution Basics

Alternating current electricity fed into the grid can generally be categorized into generation, transmission, and distribution (NERC, 2013). Generation through some form of energy (coal, gas, nuclear, etc.) occurs, and electricity is then transported across a series of high-voltage transmission and lower-voltage distribution lines to reach approximately 334 million people in homes and businesses in North America (NERC, 2013). A key component of the grid are transformers that step up electric voltage at generating stations for efficient transport and then step down voltage at substations to safely deliver low voltage electricity to customers (NERC, 2013).

Electricity flows through substations via bus to a circuit breaker that connects transmissions lines to a transformer and/or a generator to facilitate electricity distribution. Substations are connected via transmission lines, which as mentioned

above, contain the machinery and ability to step up and step down power to facilitate transmission and distribution (NERC, 2013). Throughout the twentieth and twenty-first centuries, various components of substations have become networked to facilitate operations, and thus the grid can be thought of as containing both IT components that interact with OT components.

2.2 Information Technology and Operational Technology in ICS

Major issues between the IT and OT subsets of ICS involve a lack of common terms and understanding of components (and their associated personnel), and also the end goals of what each seeks to preserve. These end goals could be simplified for the OT side as focused on reliability, and for the IT side as focused on security (Murphy, 2017). While these two sides are not diametrically opposed in most cases, something such as a patching of an operator interface during peak hours would potentially be considered unacceptable from the OT perspective, as it is inherent (at least in the norms regarding essential services in the US) that power must be constant and reliable. Given the great reliance on devices powered by traditional forms of power generation (i.e., the grid as opposed to solar), this could be catastrophic to businesses and local economics.

This focus is best exemplified by the two Reliability Concepts put forth by the North American Electric Reliability Corporation (NERC). The first is adequacy, which is

defined as “the ability of the electricity system to supply the aggregate electrical demand and energy requirements of the end-use customers at all times, taking into account scheduled and reasonably expected unscheduled outages of system elements” (NERC, 2013). The second is operating reliability, defined as “the ability of the bulk-power system to withstand sudden disturbances, such as electricity short circuits or unanticipated loss of system elements from credible contingencies, while avoiding uncontrolled cascading blackouts or damage to equipment”, in which the bulk-power system refers to all electric generation and transmission components and their associated control systems (NERC, 2013). Note, that the emphasis seems skewed to supply demand of electricity at all times, and that there is nothing regarding security of the system or any type of risk mitigation regarding shutdowns due to physical/cybersecurity concerns.

The interaction between IT and OT components in modern ICS is ubiquitous and necessary to provide essential services to the population. However, this interconnectivity produces vulnerabilities in ICS network architecture that can be exploited. This will be further explored in Section 2.5 which describes the Aurora vulnerability and Section 3.1 which provides information on the testbed architecture used to produce the dataset utilized in this research.

2.3 Synchrophasors and ICS

As the vast majority of features/attributes from the dataset are synchrophasor or Phasor Measurement Unit (PMU) measurements, the following section will provide more information about how the technology works, the case for utilizing synchrophasor technology, and what the measurements mean.

Electrical transmission systems composed of lines, breakers, and transformers form the basis for transmitting electricity from generators across long distances to load centers (Pan, Morris, Adhikari, March 2015). With the increasingly connected nature of IT and OT components within ICS infrastructure, synchrophasor technology such as Phasor Measurement Units (PMUs) are being utilized to serve as sensors to various processes occurring during transmission (Pan et al., March 2015). The sensors are capable of monitoring real-time magnitude and phase of voltage, magnitude and phase of currents, and frequency of the system in a synchronized manner (Crappe, 2008). This synchronization allows an evaluation of the stability of power system network, load distribution calculations, and fault detections/locations (Crappe, 2008). The PMU data is time stamped with Universal Time Coordinated (UTC) via Global Positions Systems signals, and Phasor Data Concentrators (PDCs) collect said data and transmit data to a control center via Wide Area Network (WAN) (Pan et al., March 2015). The PMU measurements are aggregated through a device called a Phasor Data Concentrator (PDC), which then sends said data to OpenPDC software that sorts and processes the data for operator analysis (Morris, 2018). The overall system is called a Wide Area

Monitoring System (WAMS), and through this synchrophasor technology it is possible to measure transmission rates (comprised of multiple measurements of multiple systems) at a much higher rate than traditional Supervisory Control and Data Acquisition (SCADA) systems (30 to 120 samples per second for WAMS, as opposed to 1 sample every several seconds for SCADA) (Pan et al., March 2015). In addition to the increase of sampling rate, this method of obtaining measurements is extremely precise and exhibits very low error rates (Crappe, 2008). The dataset utilized in this research is a time series of synchrophasor measurements in which all changes/variations in a given scenario are reflected via PMU data and also reflects the IT domain via device logs in the testbed architecture (such as SNORT) (See section 3.1 for a detailed description of the dataset).

2.4 Incorporating Synchrophasor Data in Signature Based IDS

Due to the complex interactions between multiple components on both the IT and OT side, ICS and CPS (computational systems that monitor and control physical systems including but not limited to control systems, sensor-based system, and autonomous systems in an ICS), security is a complicated and unique challenge that requires interdisciplinary expertise and teamwork to properly mitigate threats (Redwood, 2016). There is often a lack of communication between IT and OT components due to restricted access and security review protocols that decrease the IT side from understanding the full specifications and inherent risks of a given CPS (Murphy, 2017).

The most widely used IDS systems in CPS (and thus ICS) are Network Based IDS (NIDS), which primarily rely on signature based and anomaly based detection (Host Based IDS (HIDS) are used rarely in CPS due to resource limitations on individual ICS components and the overall complexity of CPS) (Redwood, 2016)

Signature based intrusion detection is not necessarily limited to IT components, as there are often physical indicators in OT components as to when an attack is occurring (Redwood, 2016). The benefit of utilizing an approach that focuses on sensor readings and physical measurements is ultimately that attacks can be detected regardless of the properties, stage, or attack vector (Redwood, 2016). Despite exploitation on the cyber level, many specific attacks leave indications of malicious intent in a synchrophasor through electric current events or significant changes in voltage. This has been demonstrated through research analyzing a brief power quality event in April 2015 with significant voltage sag, which was assessed and a likely source identified via PMU readings that would have been impossible to detect with legacy SCADA systems due to sampling rates (Jamei, Stewart, Peisert, Scaglione, McParland, Roberts, and McEachern, 2016).

Accurately defined signature based models formed on detection axioms/rules could be integrated into a hybrid IDS comprised of both IT and OT monitoring components and provide representations of acceptable system behavior and also detect anomalous or malicious activity. Previous research in the domain suggests that signature based rules are an integral component of a more comprehensive hybrid Synchrophasor Specific IDS (SS-IDS) comprised of both IT and OT component

monitoring (based on recommendations from the National Institute of Standards and Technology (NIST) (Khan, Albalushi, McLaughlin, Laverty, and Sezer, 2018; NIST, 2010). Of note, the research in this work is specifically focused on phasor measurement values as opposed to other features that could be measured in a SS-IDS (Khan, Albalushi, McLaughlin, Laverty, and Sezer, 2018; NIST, 2010).

Jamei et al. also proposed a synchrophasor based hybrid IDS (PMU-IDS), of which rules based physical constraints are employed to draw conclusions about the state of security in various levels of the grid. This is specifically referenced in what the authors refer to as Stage 1, in which signatures of anomalies are detected in via phasor measurement variables similar to those used in this work (Jamei et al., 2016). It should be noted that because signature based rules are derived from known and analyzed instances in a dataset, that values which are attacks that are outside of the attack parameters determined by analysis will not be detected. Despite the potential limitations of the signature based model approach for a hybrid IDS, the method can establish rules that can provide a baseline for this specific type of attack which can be verified with further testing. This work provides signature based intrusion detection axioms/rules of remote tripping command injection attacks developed via statistical/ML analysis of physical synchrophasor measurements.

2.5 Historical Case Study – The Aurora Vulnerability and Remote Tripping Command Injection Attacks

In 2007, the US Department of Energy's Idaho National Laboratory conducted an experiment known as the Aurora Event or Aurora Vulnerability that displayed the vulnerabilities of generators connected to the electric grid (Zeller, February 2011). This experiment demonstrated that an attack consisting of falsified commands over compromised communication networks could have severe ramifications on the distribution of power through the exploitation of a generator (Srivastava, Ernster, Pan, 2013). By intentionally opening and closing a breaker out of synchronism, the resulting high electrical current and torque were translated to high stress on the mechanical shaft of a generator which ultimately led to its destruction (Zeller, February 2011). The exploitation of this vulnerability is referred to as a remote tripping command injection attack and will be further analyzed in Section 3.

A basic understanding of generators is necessary to display the nuances of this attack. Generators are comprised of a magnet spinning inside a tightly wound coil of wire (also referred to as a winding or turn) which produces an electrical charge and magnetic field (electromagnetism) (Barnett and Bjornsgaard, 101). If one of the wires (a conductor) moves through the magnetic field it produces electrical pressure in the wire, and the magnetic field acts a force resisting its movement (Barnett and Bjornsgaard, 101). The energy required to push the wire through the magnetic field is equal to the electric energy generated in the wire minus the energy lost in the conversion, and thus

mechanical work is converted into electricity (Barnett and Bjornsgaard, 102). The major safety feature that prevents overstress of a generator is a circuit breaker (Barnett and Bjornsgaard, 107). A circuit breaker de-energizes components in an attempt to mitigate damage should an overload, high temperature, or other faults occur (Barnett and Bjornsgaard, 107). This mechanical work and friction from rotating parts in generators is the basic principle, and the knowledge of circuit breakers relationship to this mechanical force was exploited to facilitate a successful attack.

Protective relays in a power system monitor both the generator and main network power systems and have an intentional delay which are designed to protect the system by isolating faulty parts and preventing unnecessary tripping of power components during short period transient time (Srivastava et al., 2013). The delay results in a small window where no protection is available (Srivastava et al., 2013 and Zeller, February 2011). Aurora attacks are designed to open the circuit breaker, wait for the generator to be out of synchronism, and then reclose the circuit breaker before the relay protection system identifies the anomaly (Zeller, February 2011). Through the research conducted by M. Zeller, it was determined that less than 15 of these cycles are required to launch an attack on traditional generator protection elements. This attack can be executed either locally or remotely depending on the topology of the substation communication networks (Srivastava et al., 2013).

While this type of attack is possible of being executed by a threat actor that has significant resources and capabilities, it is unlikely that threat actors of lesser resources and/or skill would be able to successful implement this type of attack (Zeller, April

2011). Also, there are varying degrees of generator (and ultimately ICS) vulnerability based on numerous risk factors (Zeller, April 2011). For example, communications protocols between the breakers and relays and the PMU can be compromised, relay communications ports can be hacked (often utilizing default passwords/usernames), and malicious programs can be embedded into the relay which initiate at a set time or condition (Zeller, April 2011). These possible vulnerabilities indicate that there are multiple attack surfaces in a given generator that could make it susceptible to this type of attack, and also demonstrate the immense complexity of safeguarding ICS architecture (Zeller, April 2011).

It should be noted that the Aurora Vulnerability can be mitigated through sound cybersecurity practices at the organizational level and proper configuration of equipment in an ICS environment (Salmon, Zeller, Guzman, Mynam, and Donolo, 2009). Proper configuration in this context refers to setting and maintaining robust generator protection schemes during both normal and faulted conditions, which for this specific type of attack includes disabling logic/protective elements preventing fast open/close operations of a breaker/relay (Salmon, Zeller, Guzman, Mynam, and Donolo, 2009). Generator protection schemes are often not enough to thwart this attack alone due to the fact that the attack is not initiated at the generator itself and is aimed at the exploitation of a node connected to the generator but not under the purview of its protection scheme (Zeller, April 2011). Additionally, possible lapses in signal processing speed or intentional design of the system by engineers to smooth the signal via filtering and keeping the system connected opens a window for attack by limiting the relay to recognize sudden

changes in the system that might indicate malicious activity (Zeller, April 2011). Although there are ways to mitigate this vulnerability, there appear to be a limited number of utility organizations that have employed updated security measures given the amount of time and effort required to update legacy systems and the focus on providing consistent services at the expense of security (Swearingen, Brunasso, Weiss, and Huber, 2013). Given the lack of incorporating mitigation factors to lessen the risk, this type of attack is still a viable and credible vulnerability that could be exploited by a well-funded/resourced threat actor.

It is also interesting to note the disclosure timeline for the Aurora Vulnerability, in which initial disclosure began in 2008 to affected domains but that all associated documents were released by DHS accidentally after a Freedom of Information Request inquiring about a non-related malware campaign called Operation Aurora in 2014 (Waltman, 2016; Murphy, 2017). Based on this, there appears to be a lack of information sharing at least to the public if not to the broader energy sector. While there are certainly security implications of sharing this type of attack with the public, there are many potentially harmful repercussions given the prevalence of generators across multiple domains/industries. Although not directly related, information sharing regarding attacks could also be discouraged by the shift of the electric power industry from regulated utilities to a more open marketplace to foster competition amongst utility companies via deregulation through the Energy Policy Act of 1992 (Barnett and Bjornsgaard, 51 and 226). There are numerous requirements of energy companies including designing facilities, attaining all permits, adhering to maintenance/operational

and repair procedures, all of which are significant endeavors (Barnett and Bjornsgaard, 226). Information sharing is likely not high on the priority list, and actually might allow for a rival company to have a competitive advantage (Barnett and Bjornsgaard, 226). While the need for information sharing for the collective security of critical infrastructure is undoubtedly required, the conditions do not make it likely. Therefore, there is a need for independent research be conducted on open source datasets to better understand and analyze existing attack data. These findings could then be published and have the potential to help multiple affected parties without removing the financial incentives or competitive advantages of said organizations.

CHAPTER 3: RESEARCH METHODOLOGY, PREVIOUS WORK ASSOCIATED WITH DATASET, AND RESULTS

This section describes the dataset and methodology utilized to analyze and classify remote tripping command injection attacks on the electrical grid given both physical measurements and traditional cyber logs. The approach below began with the utilization of statistical methods to better understand the dataset and determine if simple/straight forward statistical methods could be applied with a high degree of accuracy and facilitate the construction of detection axioms. After this statistical approach was applied without yielding successful results, a ML approach aligned with previous work associated with the original dataset was applied with success in classification of remote tripping command injection attacks.

3.1 Dataset Description

After an exhaustive search and consultation with SMEs in the domain, the author found an open source ICS attack dataset that incorporated both IT and OT attributes. This dataset was created in a joint collaboration between researchers at Mississippi State University and Oak Ridge National Laboratories and is the only dataset that could be identified that represented OT in the form of synchrophasor data, and also common IT data such as SNORT logs (Adhikari, Pan, Morris, Borges-Hink, and Beaver, 2014). Having both IT and OT components in the dataset was important as GINA excels at

analysis of data from multiple domains and is also more indicative of hybrid IDS in ICS environments which take into consideration multiple streams of data to analyze the system state (Anderson, 2018 and Redwood, 2018). The initial dataset contained fifteen sets of data in the CSV/ARFF format comprised of six groups of power system event scenarios representing natural events, no events (or normal operational conditions), and attack events (see table below; Adhikari et al., 2014).

Table 1 - Original Dataset Scenario Types (modified from Adhikari et al., 2013)

Type of Scenario	**Description**
Short Circuit Fault	A short at a various location in a power line; location indicated by percentage range (see table below for further clarification)
Line Maintenance	Normal maintenance (not attack behavior) disables one or more relays on a specific line
Attack – Remote Tripping Command Injection	Attacker sends command to relay which causes breakers to open; Aurora Vulnerability
Attack – Relay Setting Change	Attacker changes setting of distance protection scheme on relay so that said relay will not trip for a valid fault/command
Attack – Data Injection	Attacker changes values such as current, or voltage to imitate valid faults (goal is to blind operator and cause black out)
Normal Operations	Self-explanatory

From these major groups there is a total of thirty-seven scenarios, further explained by the table below (scenario numbers 31-34 were not used in the numbering convention). Scenarios 15-20 (Remote Tripping Command Injection Attacks) and 41

(Normal Operational Conditions) were isolated from the original dataset and used for the analysis in this section and future sections of this research.

Table 2 - Expanded Table of Dataset Scenarios (modified form Adhikari et al., 2013)

Scenario Number	**Description**	**Type**
1	Fault from 10-19% on Line 1	Natural
2	Fault from 20-79% on Line 1	Natural
3	Fault from 80-90% on Line 1	Natural
4	Fault from 10-19% on Line 2	Natural
5	Fault from 20-79% on Line 2	Natural
6	Fault from 80-90% on Line 2	Natural
7	Fault from 10-19% on Line 1 w/ tripping command – data injection	Attack
8	Fault from 20-79% on Line 1 w/ tripping command – data injection	Attack
9	Fault from 80-90% on Line 1 w/ tripping command – data injection	Attack
10	Fault from 10-19% on Line 2 w/ tripping command – data injection	Attack
11	Fault from 20-79% on Line 2 w/ tripping command – data injection	Attack
12	Fault from 80-90% on Line 2 w/ tripping command – data injection	Attack
13	Line 1 maintenance	Natural
14	Line 2 maintenance	Natural
15	Remote Tripping Command Injection to R1	Attack
16	Remote Tripping Command Injection to R2	Attack
17	Remote Tripping Command Injection to R3	Attack
18	Remote Tripping Command Injection to R4	Attack
19	Remote Tripping Command Injection to R1 and R2	Attack
20	Remote Tripping Command Injection to R3 and R4	Attack
21	Fault from 10-19% on Line 1 with R1 disabled and fault – relay setting change	Attack
22	Fault from 20-90% on Line 1 with R1 disabled and fault – relay setting change	Attack
23	Fault from 10-49% on Line 1 with R2 disabled and fault – relay setting change	Attack
24	Fault from 50-79% on Line 1 with R2 disabled and fault – relay setting change	Attack

Scenario Number	Description	Type
25	Fault from 80-90% on Line 1 with R2 disabled and fault – relay setting change	Attack
26	Fault from 10-19% on Line 2 with R3 disabled and fault – relay setting change	Attack
27	Fault from 20-49% on Line 2 with R3 disabled and fault – relay setting change	Attack
28	Fault from 50-90% on Line 2 with R3 disabled and fault – relay setting change	Attack
29	Fault from 10-79% on Line 2 with R4 disabled and fault – relay setting change	Attack
30	Fault from 80-90% on Line 2 with R4 disabled and fault – relay setting change	Attack
31	Scenario Number Not Used	N/A
32	Scenario Number Not Used	N/A
33	Scenario Number Not Used	N/A
34	Scenario Number Not Used	N/A
35	Fault from 10-49% on Line 1 with R1 and R2 disabled and fault – relay setting change	Attack
36	Fault from 50-90% on Line 1 with R1 and R2 disabled and fault – relay setting change	Attack
37	Fault from 10-49% on Line 1 with R3 and R4 disabled and fault – relay setting change	Attack
38	Fault from 50-90% on Line 1 with R3 and R4 disabled and fault – relay setting change	Attack
39	L1 maintenance with R1 and R2 disabled – relay setting change	Attack
40	L1 maintenance with R1 and R2 disabled – relay setting change	Attack
41	Normal operational load changes	Natural

The power system configuration utilized to generate the data is represented below and is referred to as a 3 bus 2 generator system. As indicated in the graphic below, the primary components consist of generators (represented by G), breakers (represented by BR), and Intelligent Electronic Devices (IEDs) (utilized to switch

breakers on/off, represented by R, controls like numbered breaker (i.e., R1 controls BR1)).

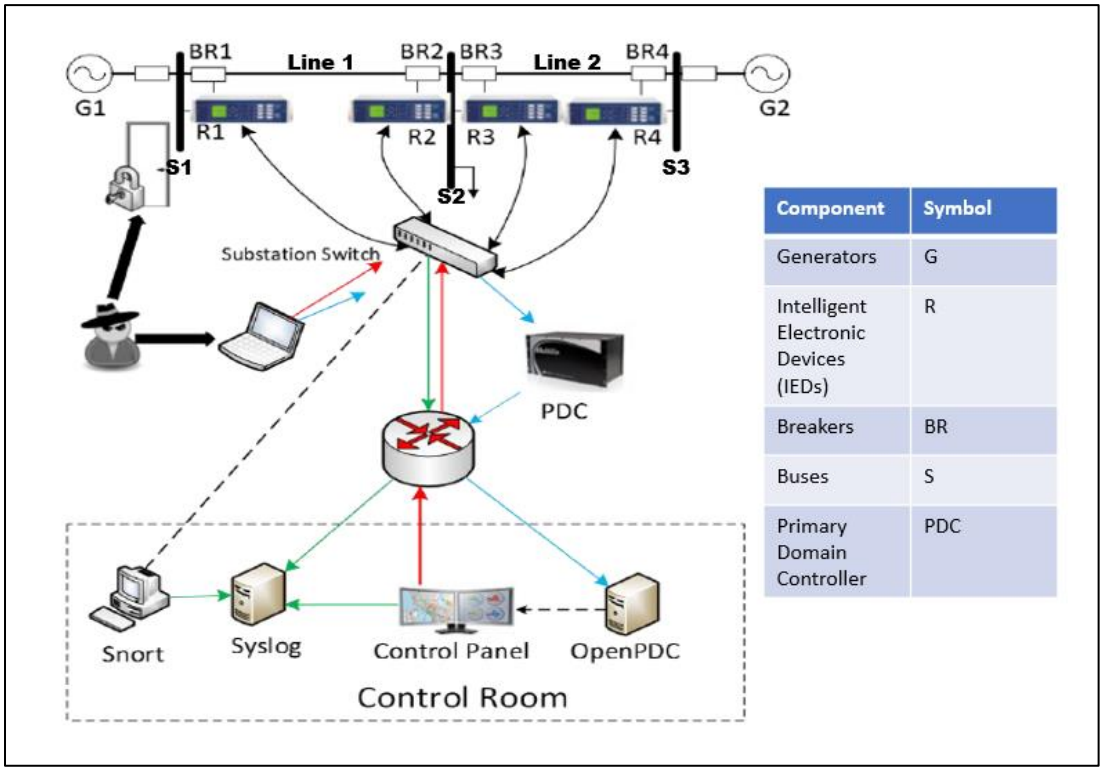


Figure 2 - Dataset Testbed Architecture (modified from Adhikarai et a., 2013)

For the original multiclass dataset, data was populated into the ARFF format, which encompassed fifteen datasets comprised of approximately 5,000 data entries each. These data are composed of 128 features or variables that are primarily sourced from phasor measurement units (PMUs) or synchrophasors. The data was measured at 120 samples per second, and each scenario was simulated for 17 seconds, which equates to each row representing 8.3 milliseconds in a CSV file (Morris, 2018).

The tables below explain the 128 measurements that comprise the dataset. In accordance with the power system configuration diagram, there are 4 PMUs, each associated with one relay which produce 29 measurements (4PMUs x 29Measurements = 116 Measurements). The remaining twelve variables are binary data associated with control panel, relay, and SNORT logs. The final column is the marker/class (see tables below): the first table shows the naming convention for all features, and the second table is a detailed naming convention for R1 features (R1 is the only relay represented for simplicity below; the same convention is used for R2-R4)).

Table 3 - Attributes/Features of Dataset (modified from Adhikari et al., 2013)

Feature (note, R1-4 will precede in raw dataset)	**Description**
PA1:VH-PA3:VH	Phase A-C Voltage Phase Angle
PM1:V-PM3:V	Phase A-C Voltage Phase Magnitude
PA4:IH-PA6IH	Phase A-C Current Phase Angle
PM4:I-PM6:I	Phase A-C Current Phase Magnitude
PA7:VH-PA9:VH	Pos.-Neg.-Zero Voltage Phase Angle
PM7:V-PM9:V	Pos.-Neg.-Zero Voltage Phase Magnitude

Feature (note, R1-4 will precede in raw dataset)	**Description**
PA10:VH-PA12:VH	Pos.-Neg.-Zero Current Phase Angle
PM10:V-PM12:V	Pos.-Neg.-Zero Current Phase Magnitude
F	Frequency for relays
DF	Frequency Delta (df/dt) for relays
PA:Z	Appearance Impedance for relays
PA:ZH	Appearance Impedance Angle for relays
S	Status Flag for relays
control_panel_log1	Self-explanatory; binary data
relay1_log	Self-explanatory; binary data
snort_log1	Self-explanatory; binary data

Table 4 - R1 Features Breakdown (modified from Adhikari et al., 2013)

Feature	**Description**
R1-PA1:VH	R1 Phase A Voltage Phase Angle
R1-PM1:V	R1 Phase A Voltage Phase Magnitude
R1-PA2:VH	R1 Phase B Voltage Phase Angle

Feature	**Description**
R1-PM2:V	R1 Phase B Voltage Phase Magnitude
R1-PA3:VH	R1 Phase C Voltage Phase Angle
R1-PM3:V	R1 Phase C Voltage Phase Magnitude
R1-PA4:IH	R1 Phase A Current Phase Angle
R1-PM4:I	R1 Phase A Current Phase Magnitude
R1-PA5:IH	R1 Phase B Current Phase Angle
R1-PM5:I	R1 Phase B Current Phase Magnitude
R1-PA6:IH	R1 Phase C Current Phase Angle
R1-PM6:I	R1 Phase C Current Phase Magnitude
R1-PA7:VH	R1 Pos. Voltage Phase Angle
R1-PM7:V	R1 Pos. Voltage Phase Magnitude
R1-PA8:VH	R1 Neg. Voltage Phase Angle
R1-PM8:V	R1 Neg. Voltage Phase Magnitude
R1-PA9:VH	R1 Zero Voltage Phase Angle
R1-PM9:V	R1 Zero Voltage Phase Magnitude
R1-PA10:VH	R1 Pos. Voltage Current Phase Angle
R1-PM10:V	R1 Pos. Voltage Current Phase Magnitude
R1-PA11:VH	R1 Neg. Voltage Current Phase Angle
R1-PM11:V	R1 Neg. Voltage Current Phase Magnitude
R1-PA12:VH	R1 Zero Voltage Current Phase Angle
R1-PM12:V	R1 Zero Voltage Current Phase Magnitude

Feature	Description
R1-F	R1 frequency for relay
R1-DF	R1 frequency delta (df/dt) for relay
R1-PA:Z	R1 appearance impedance for relay
R1-PA:ZH	R1 appearance impedance angle for relay
R1:S	R1 status flag for relay

3.2 Dataset Assumptions and Additional Information

The first major assumption of the dataset is that the IT network has been breached by an adversary (as indicated by Figure 2 above). Also of note, IEDs cannot determine if a command to open/close breakers has been issued from an adversary due to a lack of internal validation (Adhikari et al., 2014). Therefore, opening/closing breakers will not be detected as malicious as there is not a mechanism to determine where the command is coming from (i.e., from an operator or from an adversary) (Adhikari et al., 2014). In the testbed architecture, SNORT was monitoring only whether a packet had been sent to trip, and a single packet could be a legitimate command if observed by both SNORT and the overall Energy Management System from an operator (Morris, 2018). If only SNORT observed the packet however, this would indicate an attack had been initiated by an adversary (Morris, 2018). Additionally, to provide detailed and specific analysis of remote tripping command injection attacks,

only remote tripping command injection attack data and normal operational data was considered during analysis (see Section 3.4 for modification details).

3.3 Previous Work Associated with Dataset

In the next section regarding work associated with this dataset, the article that is most directly related to this work is entitled *Machine Learning Power System and Cyber Attack Discrimination* and *Classification of Disturbances* (Borges-Hink, Beaver, Buckner, Morris, Adhikari, and Pan, 2014). In this article, the original dataset was utilized in its original form with all scenarios. The research methodology and results from Borges-Hink et al. most directly impacted the ML methodology outlined in Section 3.6.

The authors theorized in this work that ML algorithms in the open source software Waikato Environment for Knowledge Analysis (WEKA) could "leverage non-linear complex relationships between power system measurements and be able to discriminate between malicious, non-malicious, and normal disturbances" (Borges-Hink et al., 2014). Utilizing 10-fold cross validation and 90/10 train/test split, a series of ML classification algorithms were applied to the dataset including OneR, NNGe , Random Forests, Naïve Bayes, Support Vector Machines (SVM), JRipper, and Adaboost+JRipper (see table below, Borges-Hink et al., 2014).

Table 5 - ML Algorithm Descriptions (modified from Borges-Hink et al., 2014)

ML Algorithm Name	**Description**
OneR	A simple learner that evaluates each features' optimum rule and chooses best from all feature sets
NNGe	Nearest neighbor like algorithm that compares new examples to surrounding datapoints
Random Forests	Tree predictors cast a vote for most popular class/input of new instance
Naïve-Bayes	Probabilistic classifier based on Bayes theorem
Support Vector Machines	Algorithm that classifies classes based on hyper planes that maximize margin between classes
JRipper	Incremental reduced error pruning algorithm that uses a separate and conquer methodology
Adaboost	Adaptive boosting; improves performance of a base algorithm based on misclassification of previous training examples

A comparison of these learners across the dataset was illustrated via the plotting of metrics including accuracy, recall, precision, and F-Measure, and results indicated that JRipper+Adaboost had the highest accuracy across all metrics (approximately 95%) (Borges-Hink et al., 2014). The authors state the high performance of JRipper+Adaboost was likely due to its tree-based approach to rule generation and the addition of a mechanism to focus on misclassification of previous data (Borges-Hink et al., 2014). It should be noted that while classification was successful regarding differentiation between disturbances and attacks, the ML algorithms were not able to classify specific fault and attack types within each larger scenario category (Pan, Morris, and Adhikari, 2015). Also, while the research conducted by Borges-Hink et al. explores classification of the entire dataset and associated scenarios, it does not specifically address classification of individual scenarios such as specific types of attacks.

Although not directly utilizing the dataset in this work, a related article entitled *An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications* details a similar application of the same ML algorithms (with the addition of J48, a decision tree algorithm) to readings from remote terminal units (RTU) in another ICS environment, a gas pipeline system (Borges-Hink, Beaver, Buckner, Morris, Adhikari, and Pan, 2013). This work included normal operational observations and also observations of similar attacks to the remote tripping command injection attack including an Illegal Process ID Attack, in which a malicious command is sent to a Programmable

Logic Controller (PLC) to modify performance, but also included other types of command injection attacks such as manipulation of the setpoint of the pipeline pressure valve, and also command injection attacks that dealt primarily with manipulating outgoing commands in the system to acquire information (address and function scans) (Borges-Hink et al., 2013). However, it must be mentioned that the dataset structures are not similar, with instances of normal operations (28,086) and command injection attacks (257, of which only 49 are similar to a remote tripping command injection) utilized in Borges-Hink et al. being very different than the dataset utilized in this work (8,737 instances of remote tripping command injection attacks, 4,405 instances of normal operations) (Borges-Hink et al., 2013).

While this difference could have an impact on the comparative findings, the major takeaways from this work are still interesting to note. The article echoes previous findings that attacks and normal operations have higher classification rates in binary datasets than multiclass datasets (Borges-Hink et al., 2013). The work also identified the highest accuracy classifiers as nearest neighbor algorithms (NNGe) and decision tree algorithms (Random Forests) (Borges-Hink et al., 2013). The NNGe ML algorithm utilizes non-nested generalized exemplars, which are defined as examples of a dataset stored in memory that instead of being stored verbatim are merged with like examples which reduces storage in memory and thus reduces classification time, a common issue that compounds with a growing dataset and ultimately can render the exemplar database useless due to lack of memory capacity or untenable times for classification (Martin, 1995). This improvement to the nearest neighbor algorithm was proven to

increase classification performance by an average of 2.6% over standard nearest neighbor algorithms and also reduces classification time by 62% due to the reduction in exemplars (Martin, 1995). The Random Forests algorithm combines tree predictors where values depend on values of a randomly sampled vector (Breiman, 2001). This randomly sampled vector is part of a greater "forest" of trees and has the same distribution of all the trees (Breiman, 2001). The trees then "vote" for the most popular class, and the method has been shown to have lower generalization errors than other classifiers (Breiman, 2001).

As demonstrated by the differing results in the two articles above, it is likely that the specific ICS environment and type of attack highly influence the classification rates and accuracy metrics of a given ML algorithm. This could indicate that an application of differing ML approaches would yield different results based on the type of specific attack being analyzed. Therefore, subsetting existing datasets could potentially provide beneficial analysis for specific attacks that would be useful in early detection and mitigation.

3.4 Data Cleaning and Initial Analysis

Due to the large number of variables (128) and samples (13,142) in the modified dataset, the first step taken during analysis was an initial observation and subsequent

cleaning of the data, in which data cleaning can be defined as the process of transforming raw data to consistent data to facilitate analysis (De Jonge and Van Der Loo, 2013). It should be noted, prior to this step, while R is capable of reading multiple file formats including both ARFF and CSV, the author converted the original multiclass ARFF files to CSV files to facilitate modification of the dataset, and also to convert the file to a format that was easily readable for initial analysis in a program that the author had experience with (Excel (Microsoft, 1987) (for R scripts detailing the conversion process, see Appendix C). The dataset was modified to include only two scenarios listed above in Section 2 of this work (scenario 41-normal operations, and scenarios 15-20-remote tripping command injection attacks). This modification was executed to simplify the dataset and facilitate remote tripping command injection attack analysis regardless of exploitation of a single relay or multiple relays. The resulting modified dataset utilized for further analysis of remote tripping command injection attacks was comprised of 13,142 instances of which 8,737 instances were attacks, and 4,405 instances were normal operations.

After initial modification of the dataset, the next step in data cleaning was to analyze the variables in the CSV file. This was executed through both manual analysis utilizing filters in Excel, and from utilizing scripts in R. From this initial analysis, the author determined that control panel logs from R1-R4 exhibited no presence of tripping throughout the entire dataset. Additionally, the author identified that SNORT Logs for Relays 1-4 had a sum total of 8 instances in which a packet was identified as sending a trip command to a relay (twice for SNORT Log 1 identifying trip commands to R1 as

indicated by Relay Log 1, once for SNORT Log 2 to identify trip commands to R2, three times for SNORT Log 3 to identify trip commands to R3, and twice for SNORT Log 4 to identify trip commands to R4). With only 8 instances of 8737 attacks (.091%), representing 8 individual scenarios out of 105 total (7.61%), this likely indicates that SNORT (and by extension other packet sniffing intrusion detection systems) is not an integral component for identifying a remote tripping command injection attack for this dataset and that synchrophasor measurements will largely be the basis for detection/classification. SNORT and other log variables were not omitted from the dataset however, as these variables are the only variables that represent IT components, one of the unique characteristics that led to selection of this open source dataset.

Also, during initial dataset analysis 1,379 instances of data were identified that exhibited infinite values in the relay appearance impedance variables (R1.PA.Z – 399 values, R2.PA.Z – 377 values, R3.PA.Z – 315 values, and R4.PA.Z – 288 values). Numerous errors occurred in processing the data in R due to these infinite values, but due to the fact that all values corresponded to attacks (that in turn corresponded to which relay was being attacked via remote tripping command injection (i.e., scenarios 15 and 19 for R1.PA.Z, scenarios 19 and 16 for R2.PA.Z, scenarios 17 and 20 for R3.PA.Z, and scenarios 18 and 20 for R4.PA.Z), it was necessary to replace the infinite values with a constant to facilitate further analysis in R.

As previous research has suggested that in both symmetric and asymmetric distributions a linear interpolation for missing data values yields high degrees of

accuracy in time series data with large sample sizes (where accuracy is measured in mean absolute percentage error, mean absolute deviation, and mean squared deviation), a modified method for imputation was utilized by taking the largest relay appearance impedance values in each original scenario and applying a multiplier of 2 to replace the infinite values in the dataset (Mahmoud, Date Unknown).[1] By doubling the largest values, a spike in the data relative to each specific scenario can be easily observed. Analysis of all scenarios in the dataset which exhibited infinite values was executed via Excel sorting filters to identify the largest values. This imputation technique could be categorized as blending a single value approach as a constant was used, and a local similarity approach as a series of values within each individual scenario was utilized (Webb-Robertson, Wilberg, Matzke, Brown, Wang, McDermott, Smith, Rodland, Metz, Pounds, and Waters, 2015). Imputation was necessary to avoid eliminating over a thousand remote tripping command injection attack instances and to retain as much data as possible in the dataset for analysis and model formation.

Upon completion of this data cleaning/transformation, the CSV file was read into R for analysis of the variables (see Appendix B). Initial plots of all variables with respect to the marker variable (whether the instance was an attack or not) were created after the loading of the data (see Appendix D). While there were ranges of specific variables that indicated an attack or normal operations, initial observation of these plots did not

[1] Note, the first iteration of this data cleaning step utilized the arithmetic mean of the five greatest values in a given scenario; this however yielded the majority of values being lower than the greatest value in a cohort. Thus, to show the spike in activity, doubling the greatest value in a scenario (regardless if this value was an outlier) was utilized

yield significant results as the majority of the instances demonstrated a wide range of values regardless if the instance was an attack or normal operations.

3.5 Statistical Methods Approach

A statistical methods approach was first employed to identify if simple and easily employable methods could provide analysis and accurate classification of the dataset. The overarching goal was to simplify and reduce the trivial elements of the dataset to facilitate axiom development. The development of a small number of detection axioms or rules was integral for implementation of a conceptual model in GINA, the original direction of this work. To begin development of these rules and further analyze the dataset given the initial observations from Section 3.4, two statistical methods were utilized.

Logistic regression and by extension Stepwise Logistic Regression (SLR) was the first statistical method employed. Logistic regression is considered a part of the generalized linear model family, in which a response variable is discrete and errors do not follow normal distributions (Lindquist, Date Unknown). This method is considered appropriate in datasets such as the one used in this work where the data is binary and categorical (Lindquist, Date Unknown). SLR refers to utilizing multiple logistic regression equations to fit the data and remove variables, and RStudio incorporates both forward and backward methods (i.e., starting the model with no variables and

testing improvement in the model by adding explanatory variables, or starting with all explanatory variables and removing them sequentially until no further improvement is possible). Hybrid techniques integrating SLR have been utilized in the cybersecurity domain to improve classification in large time series malware datasets (Huda, Bawajy, Abdollahian, Islam, and Yearwood, 2016).

SLR was utilized in this work to remove variables in the dataset that had little impact on instance classification. The ultimate goal was to reduce the complexity of the dataset and thus facilitate the development of detection axioms. After fitting a logistic model to the data (with the exception of SNORT, Relays, and Control Panel logs, as SLR cannot be applied to the factor data type for explanatory variables), the stepwise function was executed in R. Akaike Information Criterion (AIC) is a widely utilized model selection criterion based off likelihood and asymptotic properties of the maximum likelihood estimator popularized by Hirotogu Akaike in his 1973 work, *Information Theory and an Extension of the Maximum Likelihood Principle*) and is used as the primary component for goodness of fit in R during SLR (Pan, 2001). AIC is defined as AIC = - 2logL(M) + 2*K, and the removal of explanatory variables that do not affect the response variable will cause decreasing AIC values and thus indicate a model is becoming more accurate. Once these variables are identified and removed, complexity of the model is also decreased, which will help facilitate the construction of detection axioms for remote tripping command injection attacks (Shtatland, Cain, and Barton, 2001; Akaike, 1973).

After standardizing the data, logistic regression and SLR models were created in RStudio, a data analysis software environment for the R programming language (RStudio Inc., 2011) (see Appendix E for R Scripts). The AIC of the initial SLR model based on the standardized dataset was 254,339.8.4 and was reduced to 151,248.9 by the elimination of R4.PMI2.I (IED 4 Zero Current Phase Magnitude), R1. PM11.I (IED 1 Negative Current Phase Magnitude), and R3.PM7.V (IED 3 Positive Voltage Phase Magnitude). This indicates that these variables should be removed from the model to facilitate further analysis and axiom development, as they are not explanatory values which affect the outcome as to whether an instance is an attack or normal operations. By removing these variables bias and model inaccuracy will be decreased (Bozdogan, 1987).

However, utilizing the plot function in R to obtain the Residuals Scale-Location indicated that there were 3,469 instances that did not conform to the initial logistic regression model (see plot below; 9,673 instances did conform to the model as indicated by the thick line running parallel to the X axis) (Wiegand, 2018). Residuals can be thought of as False Negatives (FN), and that the models below would misclassify these instances as normal operations (Wiegand, 2018). Due to the severity of electrical outages and potential ramifications, it is necessary to catch all instances of attacks with the lowest FN rate possible.

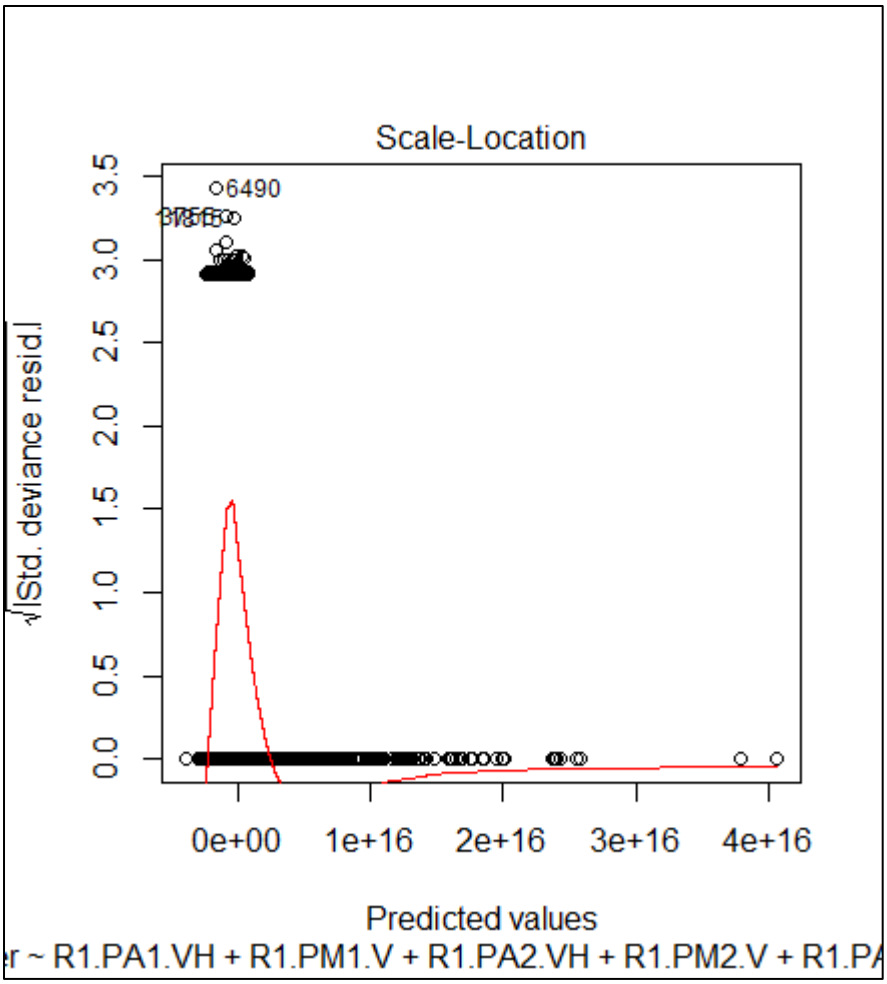


Figure 3 - Residual Plot of Initial Logistic Regression Model

Through the use of filter commands in R the class of the high residuals were identified, in which high residuals have over a value of 2.0 for the square root of the standard deviance of residual value in the plot above. There were 3,469 high residuals identified, and through the use of R Scripts all 3,469 instances were identified as of the attack class (see Appendix E for R scripts/outputs) (Wiegand, 2018). After the removal of R4.PM12.I, R1.PM11.I, and R3.PM7.V through utilization of the step function, the

concentration of residuals remained in the same location, but the number decreased (see figures below). Through a utilization of filter commands in R, there were 1,511 high residuals identified in the SLR model, and again all of these instances were of the attack class (see second figure below and Appendix E for R scripts/outputs).

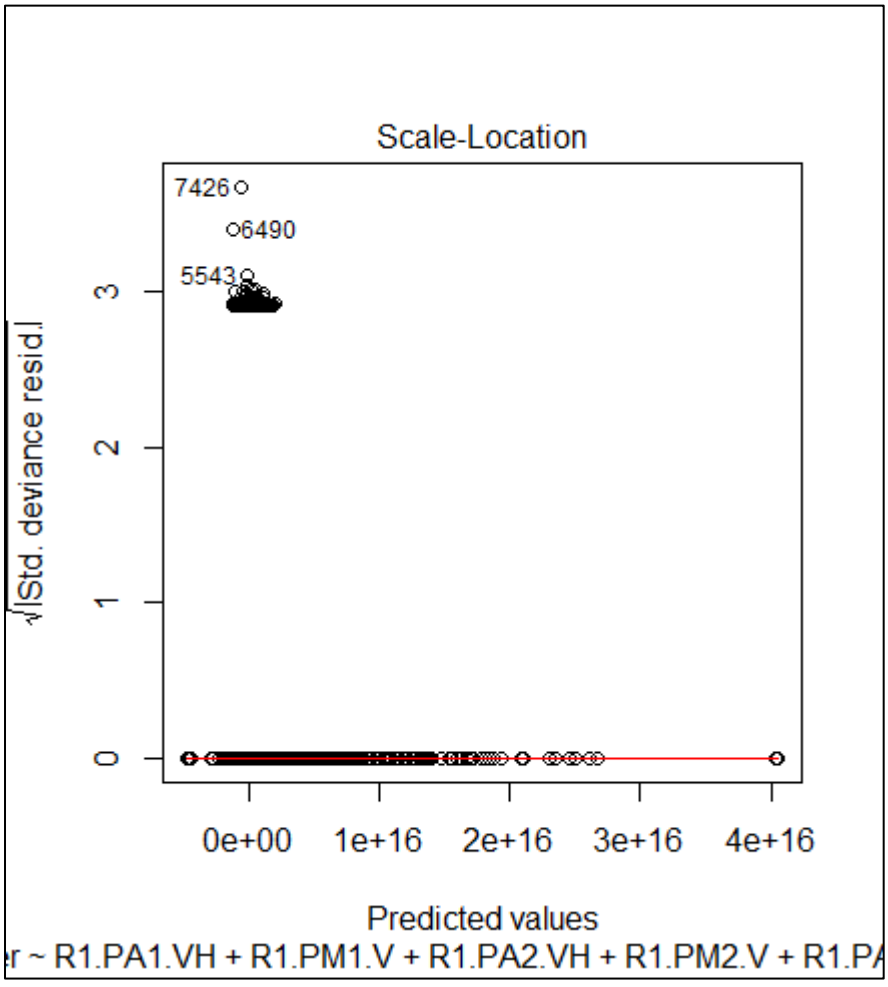


Figure 4 - Residual Plot of Stepwise Logistic Regression Model (vars R4.PM12.I, R1.PM11.I, and R3.PM7.V removed)

Table 6 – Initial Logistic and Stepwise Logistic Residual Comparison/AIC Values

Model Name	Dataset	AIC	Total Residuals	Hi-Res. Attack	Hi-Res. Normal	Low-Res. Attack	Low-Res. Normal
Standardizedmodel1 (glm)	Standardized. Aurora	254339.8	13,142	3469	0	5268	4405
Stepwisestdmodel1 (stepped Standardizedmodel1; removed R4.PM12.I, R1.PM11.I, R3.PM7.V)	Standardized. Aurora	151248.9	13,142	1511	0	7226	4405

These results indicate that while there is a subset of the dataset that can be classified, that there is a significant portion which cannot. Unfortunately, this portion of the dataset that cannot be classified easily is comprised of the attack data which this work seeks to analyze. Because the residuals indicate a high number of FNs and possible non-linearity, the dataset was split into two subsets for further examination. These subsets were named “easy” and “hard”, in which the “easy” is termed as such as the data initially seems to follow a linear pattern and is comprised of both attacks and normal operations for the response variable. The “hard” subset is termed as such as the attack data appears to not follow a linear trend as represented by the residual graph above. Additional SLR models were fitted to each of these data subsets given the residual data in an attempt to find if all variables were explanatory for these instances of attacks. The focus for further analysis was the easy subset, as the hard subset was comprised of all attacks and thus SLR would remove all variables from the initial model and the AIC value would decrease to 0 (see image below for visual representation of subsets) (Wiegand, 2018).

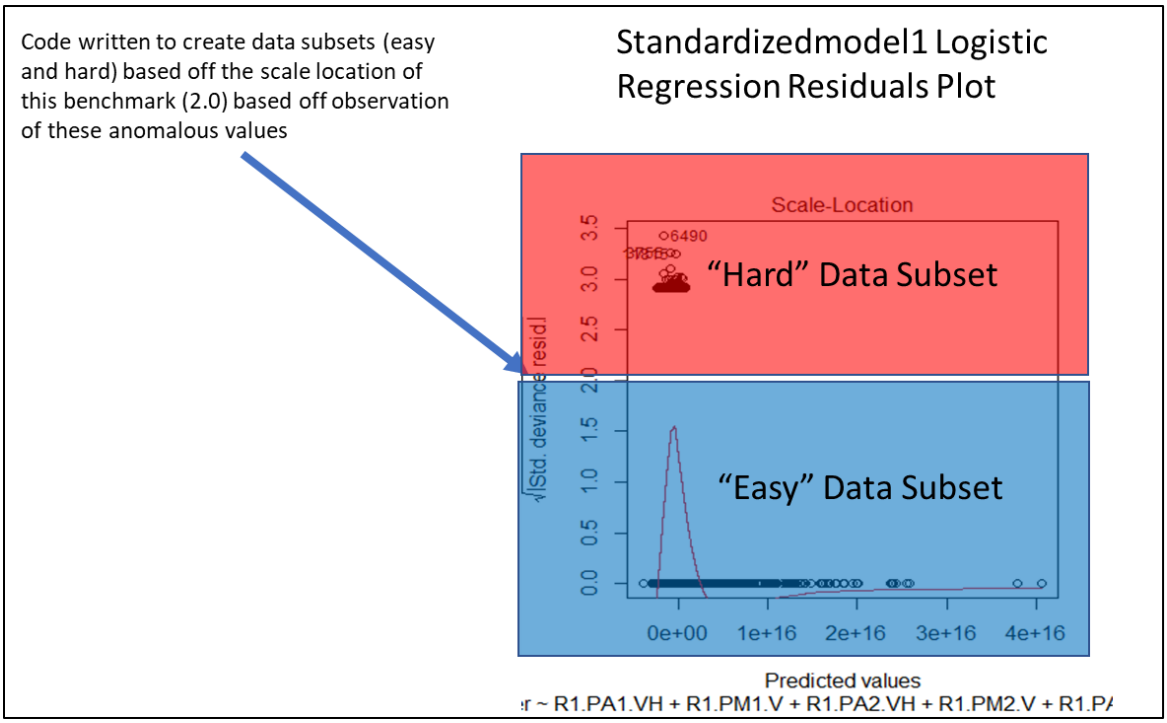


Figure 5 - Visual Representation of Data Subsets

Through plotting the residuals utilizing the same methods as the initial SLR model, it was observed that while the residual position spatially is similar in the easy data subset logistic regression model, the residual number are far less, but again consisting of all attacks (see plot and table below) (for R scripts/output, see Appendix F for R).

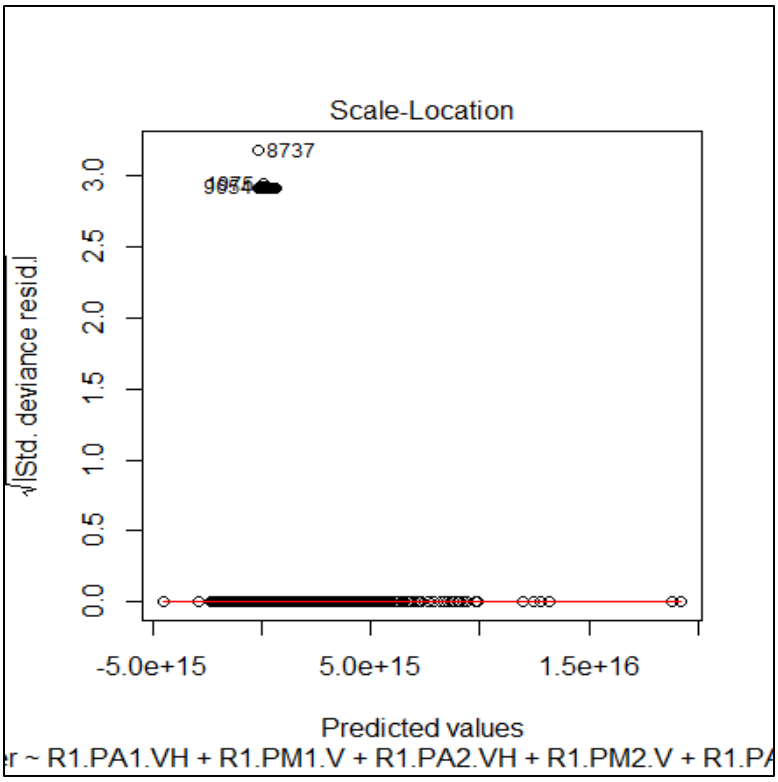


Figure 6 - Residual Plot of Easy Data Subset Initial Logistic Regression Model

After obtaining the initial easy logistic regression model from the easy data subset, the step function was utilized to perform SLR. Far lower AIC values were observed initially (5998.98), and significant AIC reduction was achieved through the removal of R2.PA4.IH (AIC reduced to 4338.98). However, plotting the residuals of this subsequent model and utilizing R Scripts to identify the numbers, it was observed that while the AIC value decreased, the number of residuals increased. Due to the implications of a FNs in the ICS context, despite lowering AIC values even one additional instance should be grounds to nullify the removal of any variable (see figures below).

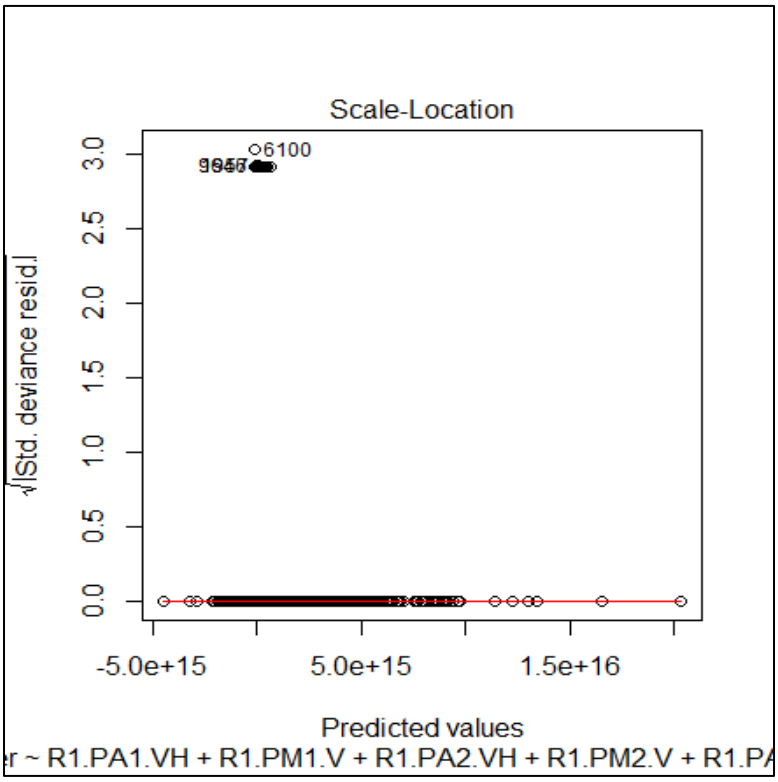


Figure 7 - Residual Plot of Easy Data Subset Stepwise Logistic Regression Model (R2.PA4.IH removed)

Table 7 - Easy Data Subset Residuals Comparison

Model Name	Dataset	AIC	Total Residuals	Hi-Res. Attack	Hi-Res. Normal	Low-Res. Attack	Low-Res. Normal
Easystdstepwisemodel1	Easy Standardized Aurora	5998.98	9673	9	0	5259	4405
Easystdstepwisemodel2 (stepped ESTSWDM1; variable removed – R2.PA4.IH)	Easy Standardized Aurora	4338.98	9673	10	0	5258	4405

Another iteration of stepwise logistic regression was executed given the easy subset of data with the removal of suggested variables from the initial SLR model

(R4.PM12.I, R1.PM11.I, and R3.PM7.V). As indicated by the tables and figures below, there were 132 residuals present in the initial logistic regression model. After application of the step function, the AIC decreased from 10462.4 to 4615.3, and the residual number decreased through the removal of R2.PA2.VH, R1.PM8.V, R4.PA12.IH, an R3.PA9.VH (see figures/tables below).

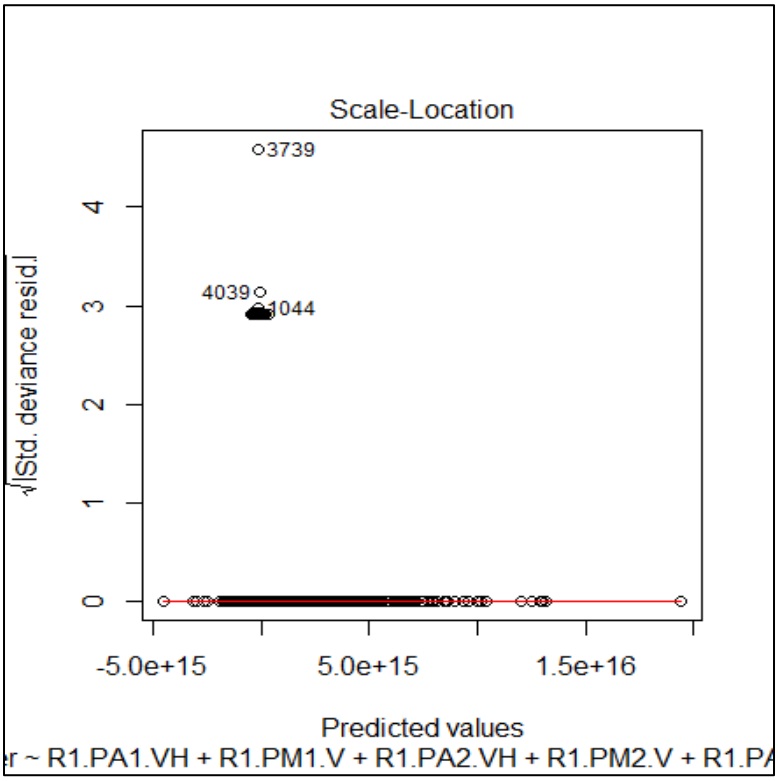


Figure 8 - Easy Subset Logistic Regression Model 3 (with removal of R4.PM12.I, R1.PM11.I, and R3.PM7.V)

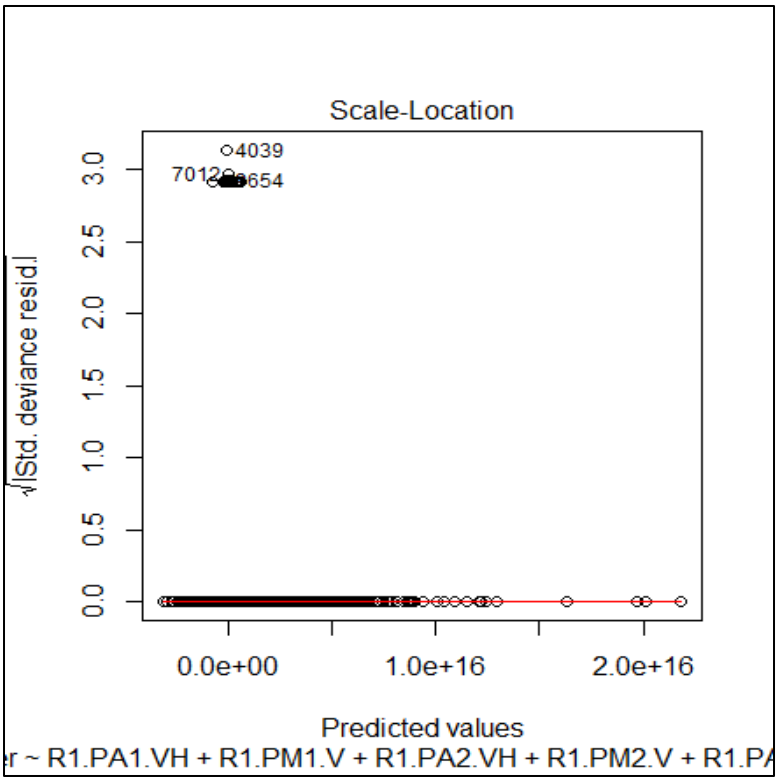


Figure 9 - Easy Subset Stepwise Logistic Regression Model 4 (with removal of R2.PA2.VH, R1.PM8.V, R4.PA12.IH, and R3.PA9.VH)

Table 8 - Easy Subset Residuals Comparison (with omission of initial logistic regression variables)

Model Name	Dataset	AIC	Total Residuals	Hi-Res. Attack	Hi-Res. Normal	Low-Res. Attack	Low-Res. Normal
Easystepwisemodel3 (removed R2.F, R1.PM11.I, R2.PM11.I, and R1.PA2.VH from initial stepwise model)	Aurora.easy	4689.41	9554 (50 high, 9504 low)	50	0	5099	4405
Easystepwisemodel4 (stepped EM3; variable removed - R2.PA4.IH)	Aurora.easy	2524.79	9554 (6 high, 9548 low)	6	0	5143	4405

In summary, SLR yielded minimal results due to complexity of the dataset and large degrees of non-linearity indicated by residuals plots. More evidence of this lies in

that of the three iterations of applying a step function to the logistic regression model, while the AIC did decrease (at vastly different rates), the removal of variables was inconsistent across all iterations. Through the utilization of SLR, eight unique variables were identified to be omitted based on differing configurations of the initial logistic regression models. Additionally, perhaps the most telling indicator of complexity and non-linearity which creates difficulty in classification in the dataset is that all residuals are attacks. The residual rate indicates thousands of misclassifications and less than a 90% accuracy classification, which is not an acceptable metric given the gravity and implications of an attack on the grid. To further analyze the dataset, another statistical method was employed.

The second statistical method utilized was Principal Component Analysis (PCA). This method was formulated by Karl Pearson and is often regarded as forming the basis for multivariate data analysis (Wold, Esbensen, Geladi, 1987). PCA is conducted through the approximation of a larger matrix via the product of two smaller matrices (Wold et al., 1987). This analysis was designed in Pearson's words as trying to find "lines and planes of closest fit to systems of points in space", and the major goals of the method include simplification of data, data reduction, general modeling, outlier detection, variable selection, classification, prediction, and unmixing of data (Wold et al., 1987). The majority of these goals are concerned with simplifying data for analysis, which is aligned with the primary objective of this works initial statistical analysis (Wold et al., 1987). The hope was that utilizing this method could simplify massive amounts of

complex information into an understandable, palatable, and potentially actionable size for axiom or rule development.

PCA was also primarily executed via formulation/execution of scripts in RStudio (RStudio Inc., 2011). The dataset was first standardized and all non-factor variables were removed and the prcomp function was utilized to create a model of the data utilizing PCA (see Appendix A) (Wiegand, 2018; Coghlan, 2013). To visualize the data, the screeplot function was used, and an analysis revealed that at approximately the 24th to 25th transformation the slope starts to level off, which indicates a high degree of variance in the data (see image below) (Wiegand, 2018).

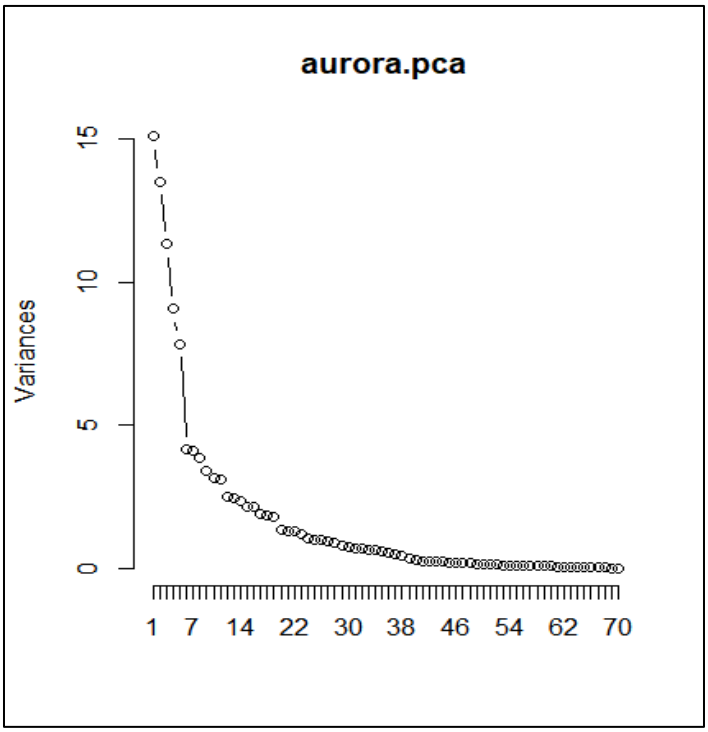


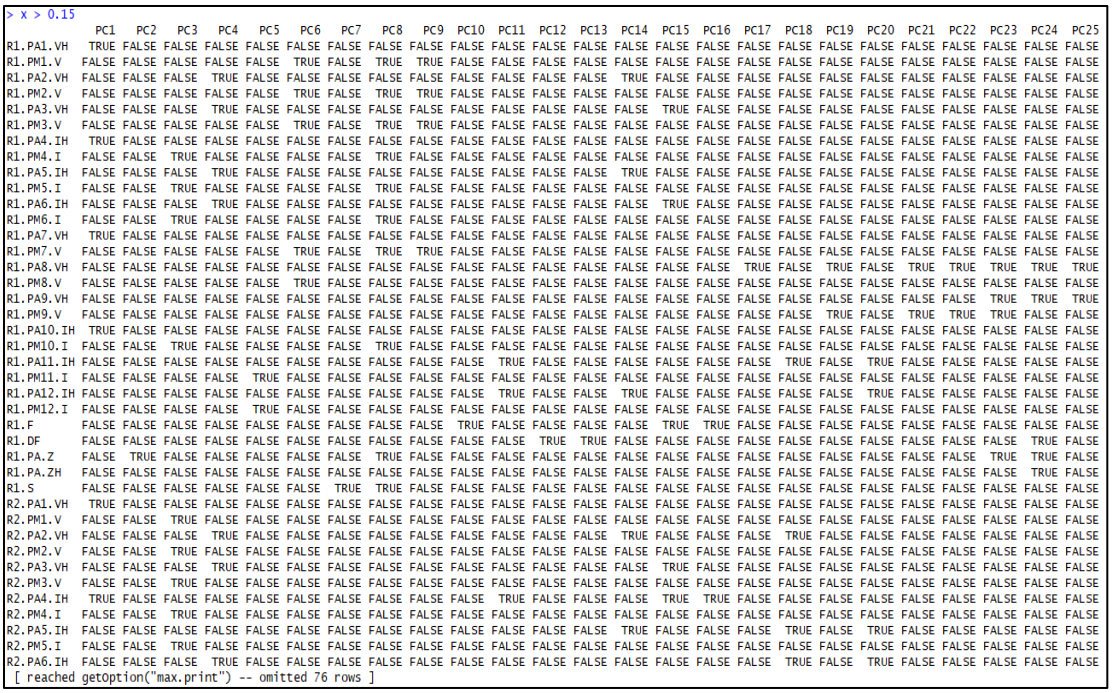
Figure 10 - PCA Scree Plot

This was further confirmed through the application of Kaiser's Criterion which as applied to the standardized data would include all translations where the variance was greater than 1, which was through the 25th transformation (see figure below).

```
> aurora.pca <- prcomp(standardizeAll)
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7
Standard deviation     3.884 3.6745 3.3648 3.00857 2.79950 2.03728 2.02047
Proportion of Variance 0.130 0.1164 0.0976 0.07803 0.06756 0.03578 0.03519
Cumulative Proportion  0.130 0.2464 0.3440 0.42206 0.48962 0.52540 0.56060
                           PC8     PC9   PC10    PC11    PC12    PC13    PC14
Standard deviation     1.96414 1.85016 1.7794 1.76289 1.57755 1.56727 1.53027
Proportion of Variance 0.03326 0.02951 0.0273 0.02679 0.02145 0.02118 0.02019
Cumulative Proportion  0.59385 0.62336 0.6507 0.67745 0.69890 0.72008 0.74027
                          PC15    PC16    PC17   PC18    PC19   PC20   PC21
Standard deviation     1.47435 1.46207 1.37590 1.3582 1.34295 1.1652 1.1399
Proportion of Variance 0.01874 0.01843 0.01632 0.0159 0.01555 0.0117 0.0112
Cumulative Proportion  0.75901 0.77743 0.79375 0.8097 0.82520 0.8369 0.8481
                          PC22    PC23    PC24    PC25    PC26    PC27    PC28
Standard deviation     1.13555 1.08498 1.03020 1.00538 0.99020 0.98280 0.94094
Proportion of Variance 0.01112 0.01015 0.00915 0.00871 0.00845 0.00833 0.00763
Cumulative Proportion  0.85923 0.86937 0.87852 0.88724 0.89569 0.90402 0.91165
```

Figure 11 - PCA Standard Deviation and Proportion of Variance

Through an examination of the coefficients in the 25 rotations of the principal components, in each of the 116 variables the absolute value of the coefficient was greater than .15 in at least one of the principal components, meaning all the variables were needed in at least one rotation, but many were needed in multiple rotations (see truncated figure below, where x is a variable that represents the absolute value of a principal component rotation given a variable).



```
> x > 0.15
             PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10  PC11  PC12  PC13  PC14  PC15  PC16  PC17  PC18  PC19  PC20  PC21  PC22  PC23  PC24  PC25
R1.PA1.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM1.V   FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA2.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM2.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA3.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM3.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA4.IH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM4.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA5.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM5.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA6.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM6.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA7.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM7.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA8.VH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
R1.PM8.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA9.VH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
R1.PM9.V   FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE
R1.PA10.IH  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM10.I  FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA11.IH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
R1.PM11.I  FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA12.IH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM12.I  FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.F       FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.DF      FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
R1.PA.Z    FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
R1.PA.ZH   FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
R1.S       FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA1.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM1.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA2.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM2.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA3.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM3.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA4.IH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM4.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA5.IH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
R2.PM5.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA6.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
 [ reached getOption("max.print") -- omitted 76 rows ]
```

Figure 12 - PCA Coefficient Output

The results of PCA further suggest that the dataset is non-linear and will likely be difficult to analyze via statistical methods. Also, the outputs suggest there is no simple elimination of variables and thus no simplification of the dataset based on the variance (Wiegand, 2018).

With the inability to simplify and reduce the dataset using the statistical methods of PCA and SLR, axiom development through said methods is not feasible. The failure of PCA and SLR indicates that the complexity of the dataset exceeds the capacity of statistical methods to classify the data and that simple coding of rules/axioms to classify an instance as an attack or normal operations is not possible based off this research (and the given dataset). Therefore, more sophisticated methods must be explored to

both create detection axioms for remote tripping command injection attacks and classify the dataset with a high degree of accuracy.

3.6 Machine Learning Algorithm Approach

As Section 3.3 provided compelling applications of ML algorithms in previous research to classify scenarios in ICS, a similar ML methodology to the Borges-Hink et al., 2014 article was utilized to determine if results could be replicated/exceeded given remote tripping command injection attack and normal operational data. Aligned with this previous related work, the author utilized 10-fold cross validation with a 90/10 train/test set to train the ML classifiers. Notable differences from the original research include the utilization of only one dataset (binary) as opposed to the original experiment which utilized three types of datasets (multiclass, three class, and binary). Additionally, the original research conducted by Borges-Hink et al., 2014 applied the ML algorithms to a 1% random sample of the original dataset, whereas this work considers the entire modified dataset to increase data and thus analysis of remote tripping command injection attacks.

The primary tool utilized in this research to employ ML and data mining techniques was open source software called Waikato Environment for Knowledge Analysis (WEKA) (University of Waikato, 1999). The WEKA project began at the University of Waikato in 1992 with the goal of the creation of a unified workbench that

would allow researchers access to a wide collection of ML techniques/methods in one platform ((Hall, Frank, Holmes, Pfahringer, and Reutemann, 2009; Witten, Date Unknown). While multiple ML algorithms and techniques were available at the time, there was a wide array of languages/formats/platforms and no unifying application that could be utilized to easily compare and contrast differing algorithms which is necessary to determine applicability of a given method/technique (Hall et al., 2009; Witten, Date Unknown). WEKA fixed this issue by allowing users to rapidly compare different ML methods on datasets, and the software has been widely used in both academia and the private sector since its initial software release (Hall et al., 2009).

The WEKA project was initially funded by the government of New Zealand and was launched with an internal beta stage software release in 1994, and a release to the public in 1996 (Hall et al., 2009). Due to increasing complications in the software such as increasing changes to support libraries and complexity of configuration in the original C coding, the system was rewritten in Java and rereleased in 1999 (Hall et al., 2009). In addition to robust data visualization capabilities, WEKA Explorer (one off the four WEKA interfaces) was primarily utilized in this work for data classification and rule/model formulation. WEKA also includes interfaces for large scale performance comparisons for differing ML methods on differing datasets (WEKA Experimenter), a graphical interface (WEKA KnowledgeFlow), a unified interface (WEKA Workbench), and a command line interface (WEKA Simple CU) (Whitten, Date Unknown). WEKA does distinguish between data mining and ML, where data mining can be thought of as the acquisition and transformation of raw data into information that can in turn be

utilized to answer a given question or hypothesis (likely through the construction of a predictive model), and ML can be thought of as the underlaying framework which solves said question or hypothesis via the application of algorithms (Whitten, Date Unknown).

The diversity of academic disciplines that WEKA has been applied to are numerous and wide spread, indicating an adaptable and applicable tool to apply ML methods to a range of datasets and domains. In just the last year, WEKA has been used as an integral component in research focused on maximizing the utility of regression models in ML, online estimation of discrete/continuous/conditional densities, ambient sensing of detection for relay attacks in near field communication devices utilizing random forests, and even examining the intensity of emotion through the analysis of tweets on the popular social media platform Twitter (Branco, Torgo, Ribeiro, Frank, Pfahringer, and Rau, 2017; Geilke, Karwath, Frank, and Kramer, 2017; Gurulian, Shepherd, Frank, Markantonakis, Akram, and Mayes, 2017; Mohammad and Bravo-Marquez, 2017; University of Waikato, Publications Page, 2018).

WEKA has also been utilized in research associated with cybersecurity and IDS. Extensive WEKA classifier performance comparison has been applied in analysis of attack signatures given the KDD99 dataset (Nguyan and Choi, 2008; Modi and Jain, 2015). Anomaly based network intrusion detection has also been explored via the use of WEKA Naïve Bayes and Decision Tree (J48) classifiers (Nevlud, Bures, Kapicak, and Zdralek, 2013). Discrimination of malicious network communications and minimizing reliance on a human operator to interpret an insurmountable amount of data has propelled ML methods to prominence in cybersecurity, and there is utility in applying

WEKA to research focused on ICS security (Borges-Hink, Beaver, Buckner, Morris, Adhikari, and Pan, 2013). This open source software was utilized due to its noted performance in multiple domains, it's previously established related research, and also due to its user friendly/intuitive nature and a manageable learning curve.

As mentioned above, for this research the WEKA Explorer interface was primarily utilized (see figure below). After loading the dataset and applying preprocessing filters to the data to ensure the software read in the proper datatypes (see Appendices H and I for detailed steps in WEKA), all variables (which WEKA refers to as attributes) were viewed to both confirm previous data analysis in R and further visualize the data given WEKA's capabilities (see figure below in which the red bars in the histogram represent attacks and the blue represent normal operations) (Whitten, Date Unknown). Data analysis was consistent with that already executed via R scripts (i.e., numbers of attacks/normal operations). Mean, standard deviation, and the range of values differed due to use of the original non-standardized dataset, which was aligned with previous associated research. Classification utilizing ML algorithms was the next step executed in WEKA.

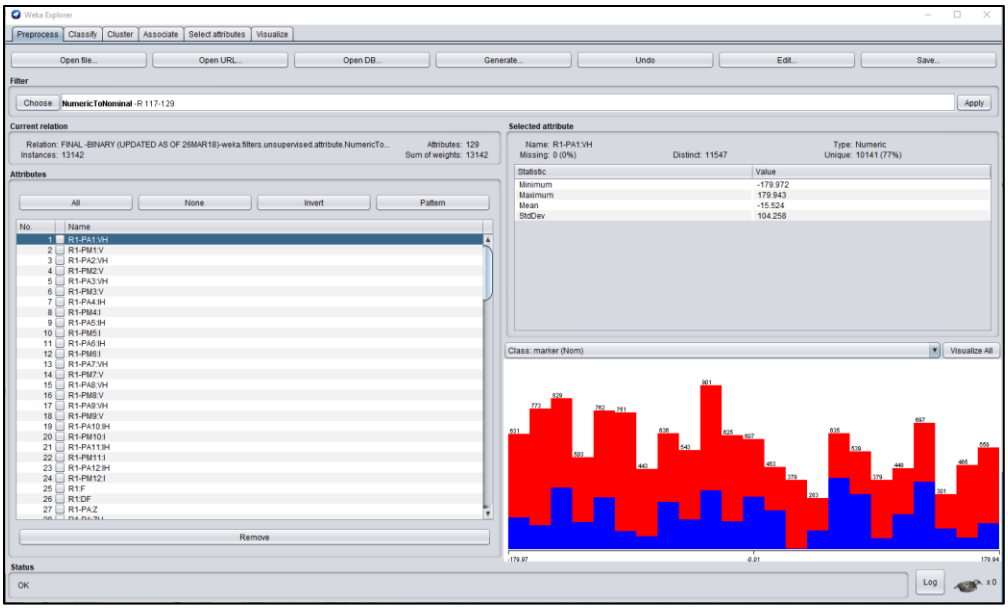


Figure 13 - WEKA Explorer Interface - Attribute Analysis

WEKA divides classification algorithms into seven distinct groups: Bayesian, functions, lazy, meta, miscellaneous, rules, and decision trees (see table below for descriptions of each). While research with the dataset in the past had tested a relatively small number of ML algorithms using the default parameters, this work tested all algorithms that could be applied to the dataset given classification specifications regarding data types. The initial number of classifiers tested was 69, which decreased to 15 based off the 95% accuracy metric established by Borges-Hink et al. and further detailed in the testing methodology below.

Table 9 - WEKA ML Classification Algorithm Groups (Brownlee, 2016; Tatsis, Tjortjis, and Tzirakis, 2013)

Classification Group	**Description**
Bayesian	Uses Bayes theorem in some capacity which predict class values by probabilities
Functions	Can be written as equation and estimates a function
Lazy	Stores training instances and work occurs during classification
Meta	Combine multiple algorithms and convert them to more powerful learners
Miscellaneous	Don't fit easily into other groups
Rules	Generates rules to classify the data
Trees	Uses decision trees based off root attributes and leaf nodes

A three-phase testing/training approach was utilized in which ML classification algorithms were compared via extraction of key metrics from the WEKA Explorer Classification panel (see figure below, as briefed to thesis advisory committee on 31MAY18).

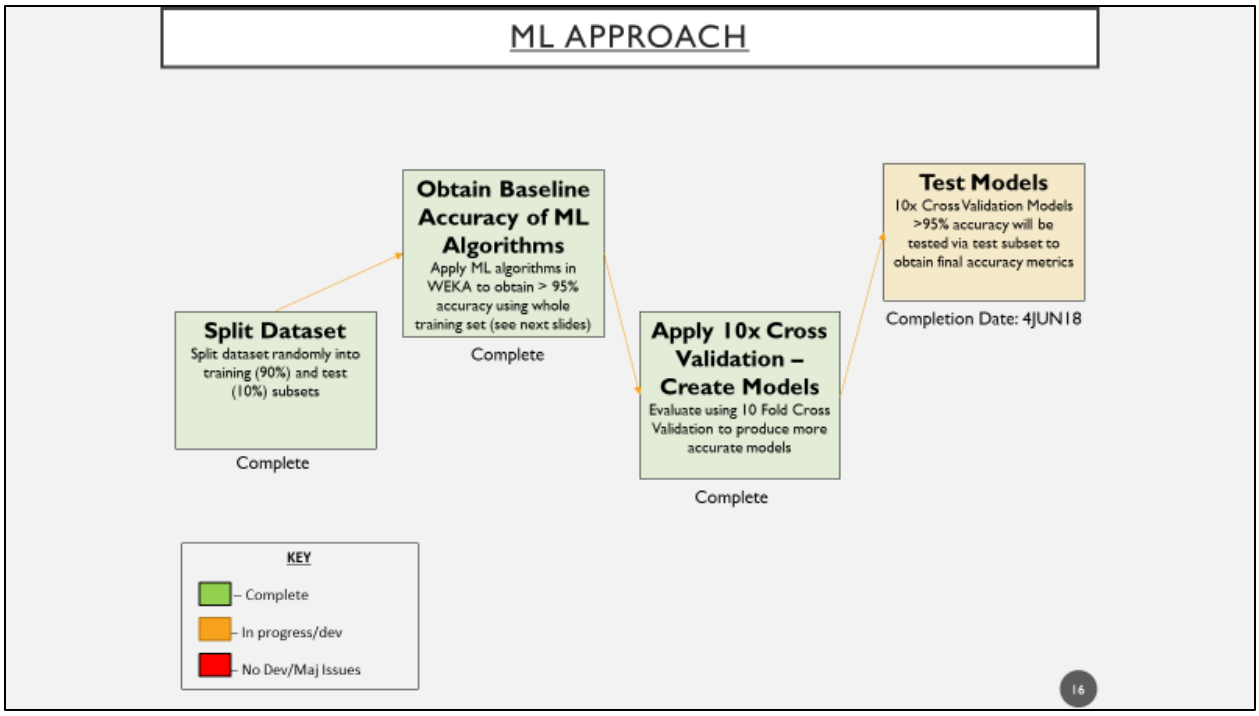


Figure 14 - ML Approach

After splitting the dataset to a 90% training set and a 10% test set using Ablebits software for randomization in the dataset CSV file, the first phase utilized the full training dataset to obtain baseline accuracy metrics and create classifier models (see figure below) (Ablebits Software, 2015).

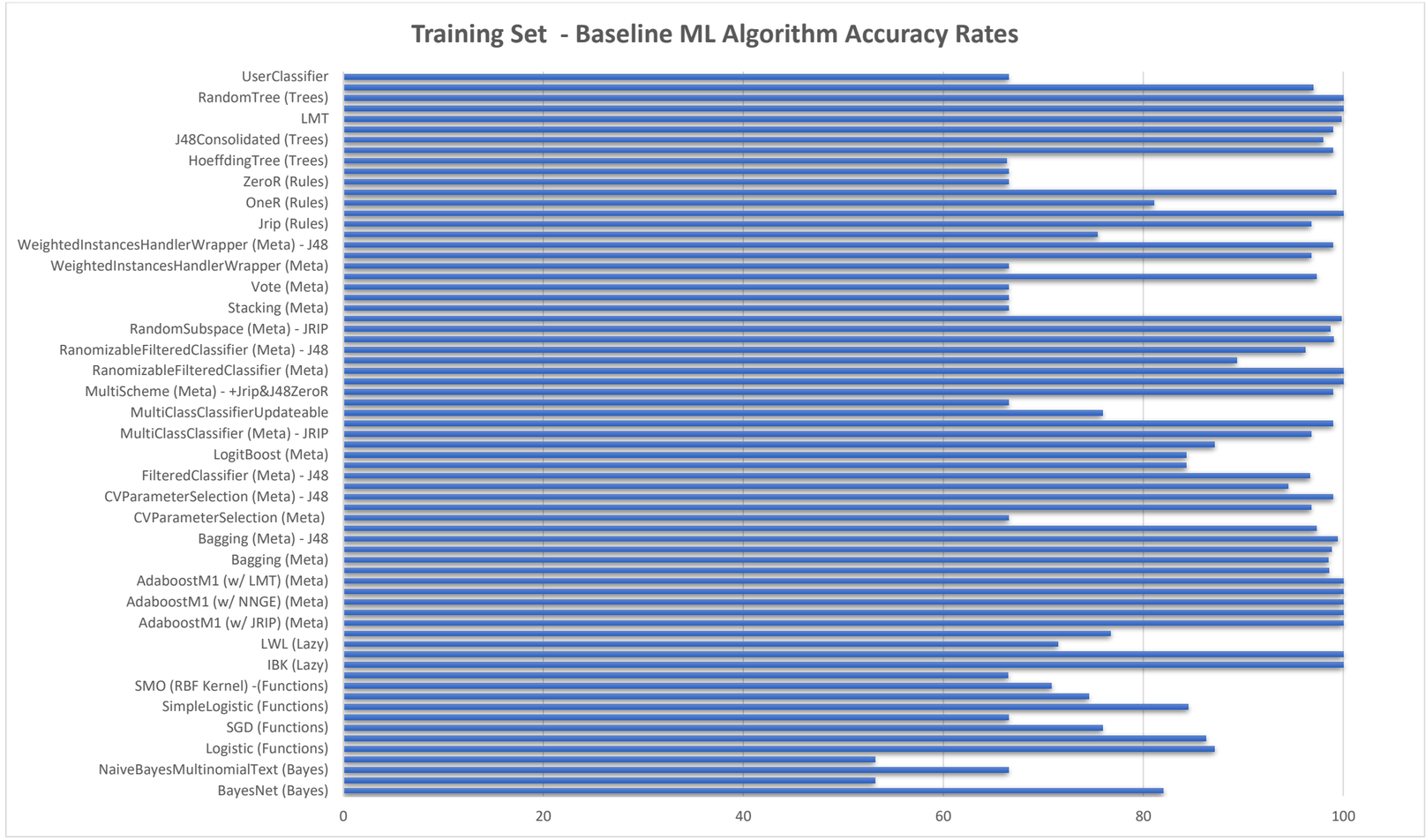


Figure 15 - Training Set - Baseline ML Algorithm Accuracy Rates

Initial results indicated thirty-seven algorithms met the 95% baseline accuracy established in previous research. The thirty-two classifiers that were omitted based on having less than a 95% accuracy rate largely consisted of simplistic rule based learners, Bayesian, and functions-based classifiers. The majority of classifiers that had 95% classification accuracy or greater consisted of meta classifiers (see figure below).

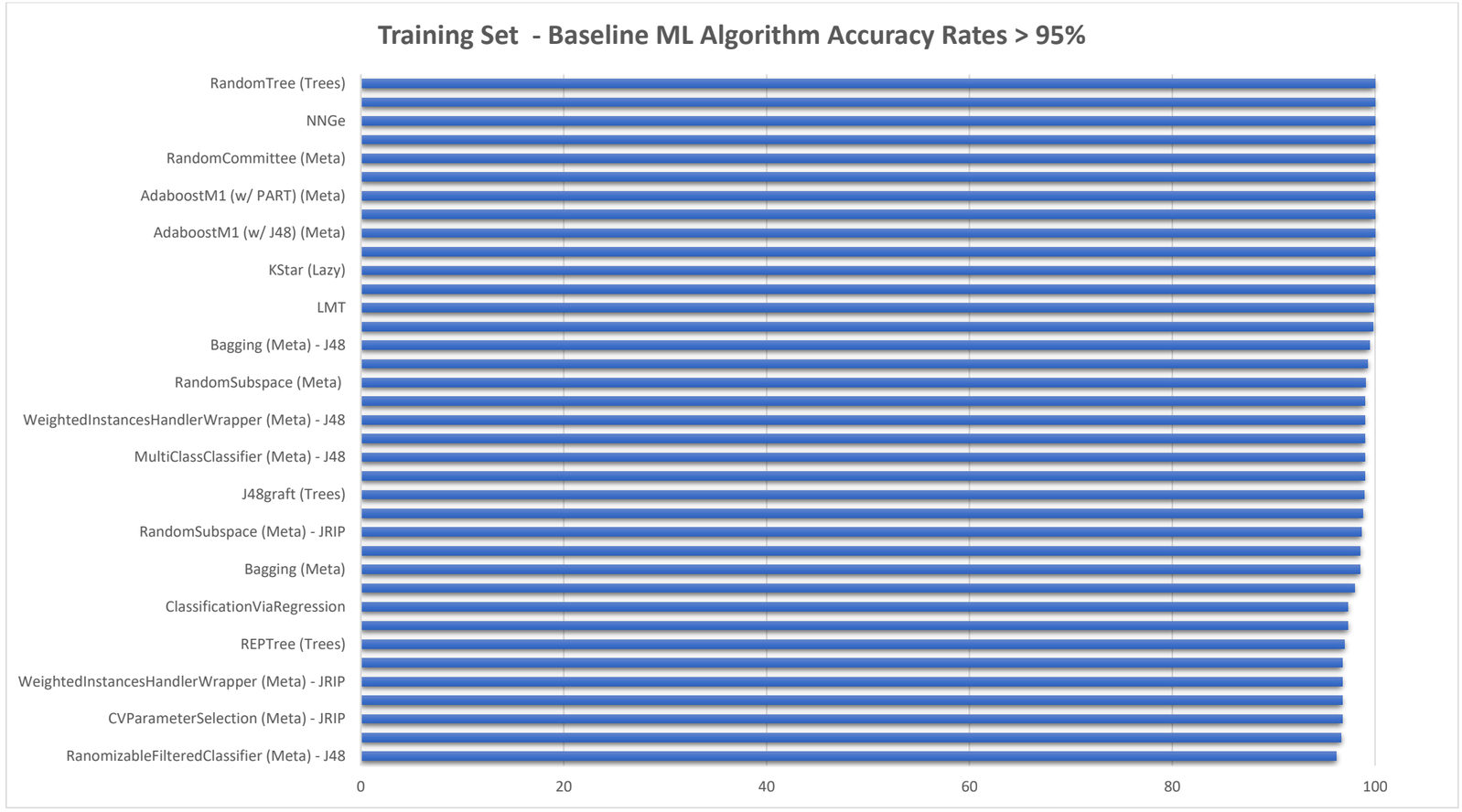


Figure 16 - Baseline ML Algorithm - Accuracy > 95%

The second phase of the ML approach utilized 10-fold cross validation to evaluate the classifier models based on the training dataset. Accuracy for all ML algorithms decreased during the evaluation utilizing 10-fold cross validation, and of the 37 algorithms that were initially included in this stage, 22 had classification rates over 95% (see figures below). Of note, the majority were again meta classifiers (15), followed by tree classifiers (5) and lazy classifiers (3).

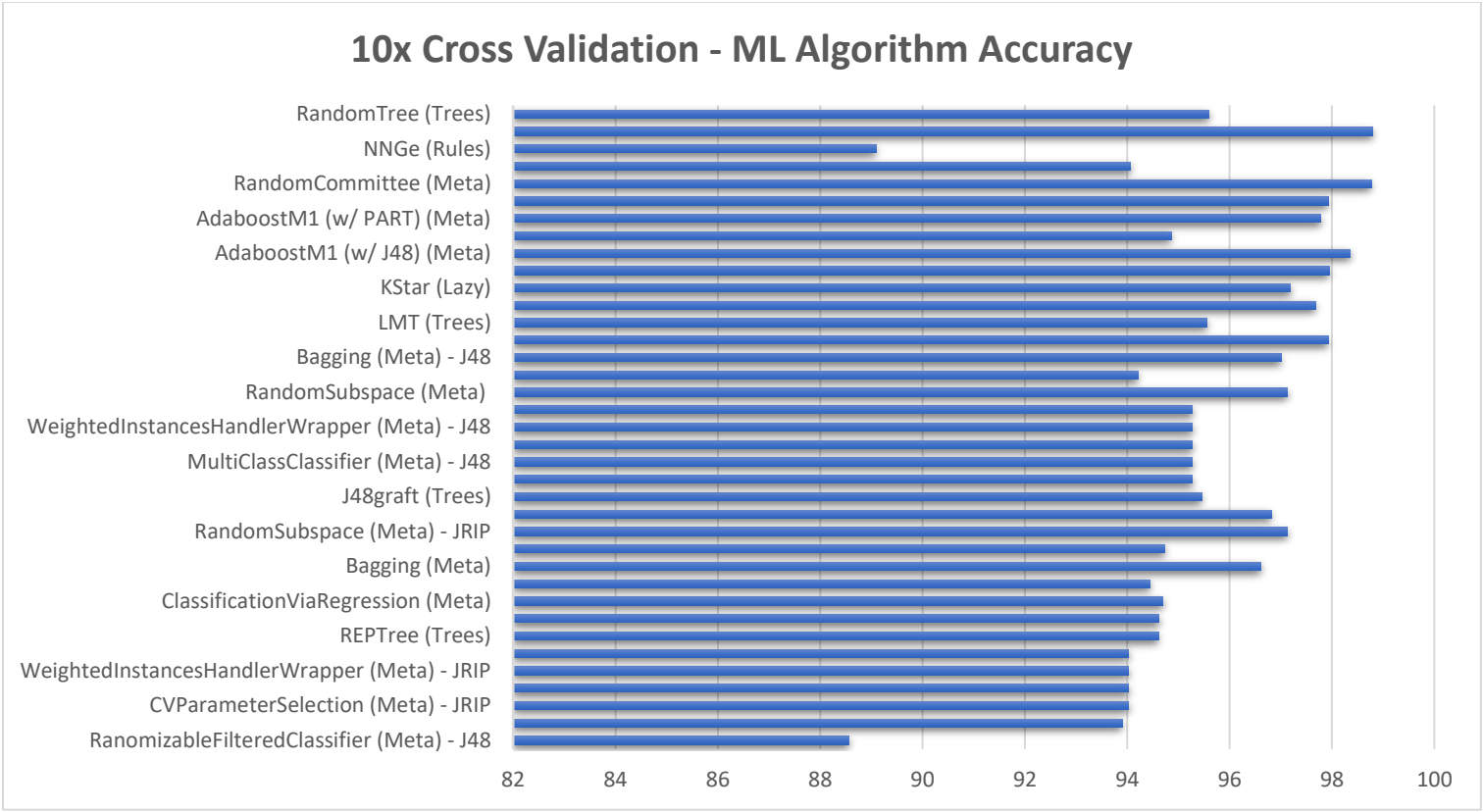


Figure 17 - 10x Cross Validation

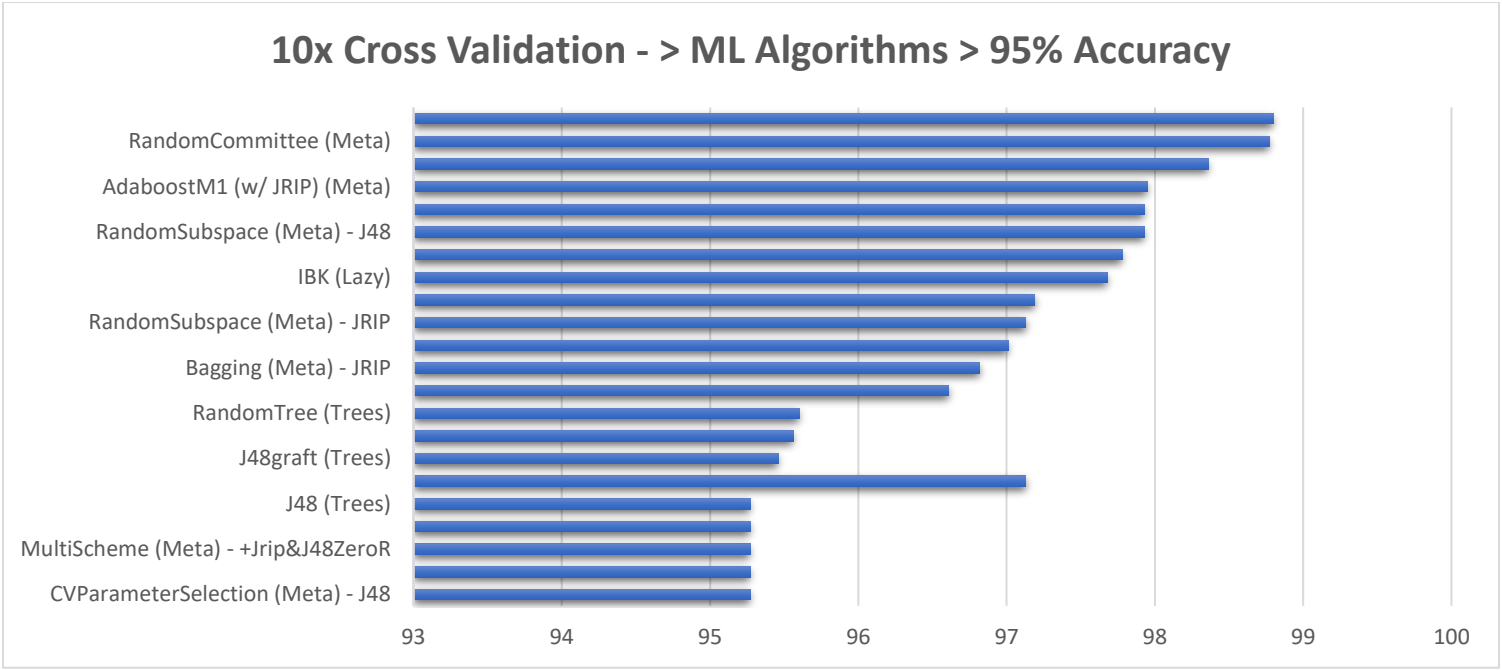


Figure 18 - 10x Cross Validation - >95% Accuracy

After identification of the algorithms that were evaluated at greater than 95%, accuracy, the models were applied to an independent dataset. To better evaluate the ML algorithms and identify the most ideal classification scheme, additional metrics captured during this final iteration included Root Mean Squared Error, FN rates for the attack class, Recall, Precision, F-Measure, and the Receiver Operating Characteristic (ROC) Curve (see table below for brief descriptions).

Table 10 - Additional ML Metrics (Holmes, 2000, Borges-Hink et al., 2014; Whitten, Date Unknown)

Metric	**Description**
Root Mean Squared Error	A standard metric for measuring the spread of y values about the predicted y value; found by squaring the residuals, averaging the squares, and taking the square root
Recall	Measures the true positive rate
Precision	Measures the positive predictive value
F-Measure	Harmonic mean of Recall and Precisions
ROC Curve	A plot that measures classification accuracy of the first class against the classification accuracy of the second

Metric	Description
	class; maximization of area under the curve indicates highest measure of classification accuracy
Time Taken	The time taken to apply the model to the test set; only utilized for comparison of top 3 models (no graphs)

The initial test set comprised of 22 ML algorithms was decreased to a final 16 algorithms, which was further decreased to 15 algorithms based off redundant results from the RandomSubspace with the default REPTree base classifier and RandomSubspace with J48 as the base classifier. Since the outputs were the same, only the RandomSubspace with the default base classifier was retained (see tables below). Of the final 15 ML algorithms with accuracy over 95% as evaluated on the test set, 10 were meta classifiers, 3 were tree classifiers, and 2 were lazy classifiers.

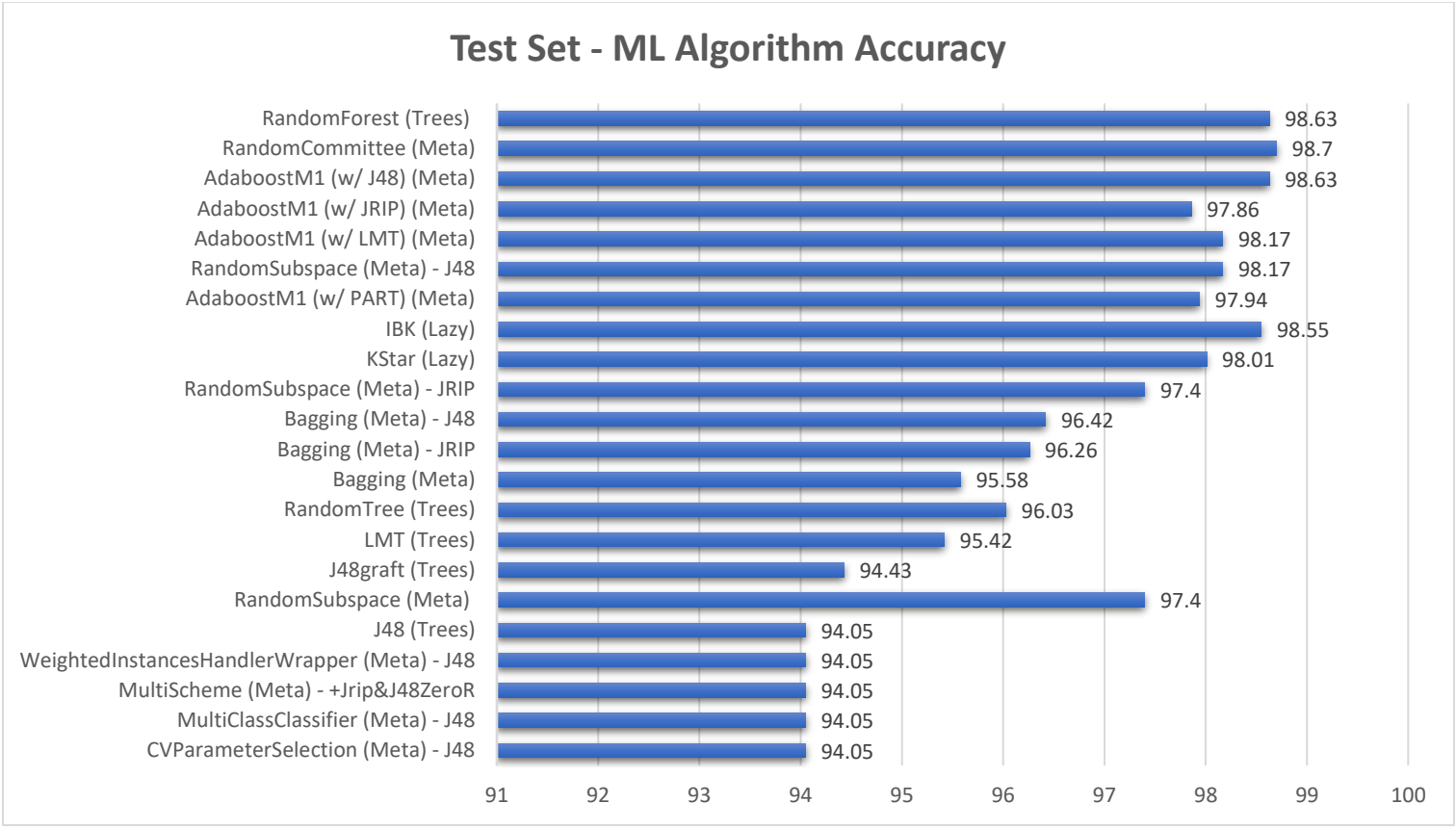


Figure 19 - Test Set - ML Algorithm Accuracy

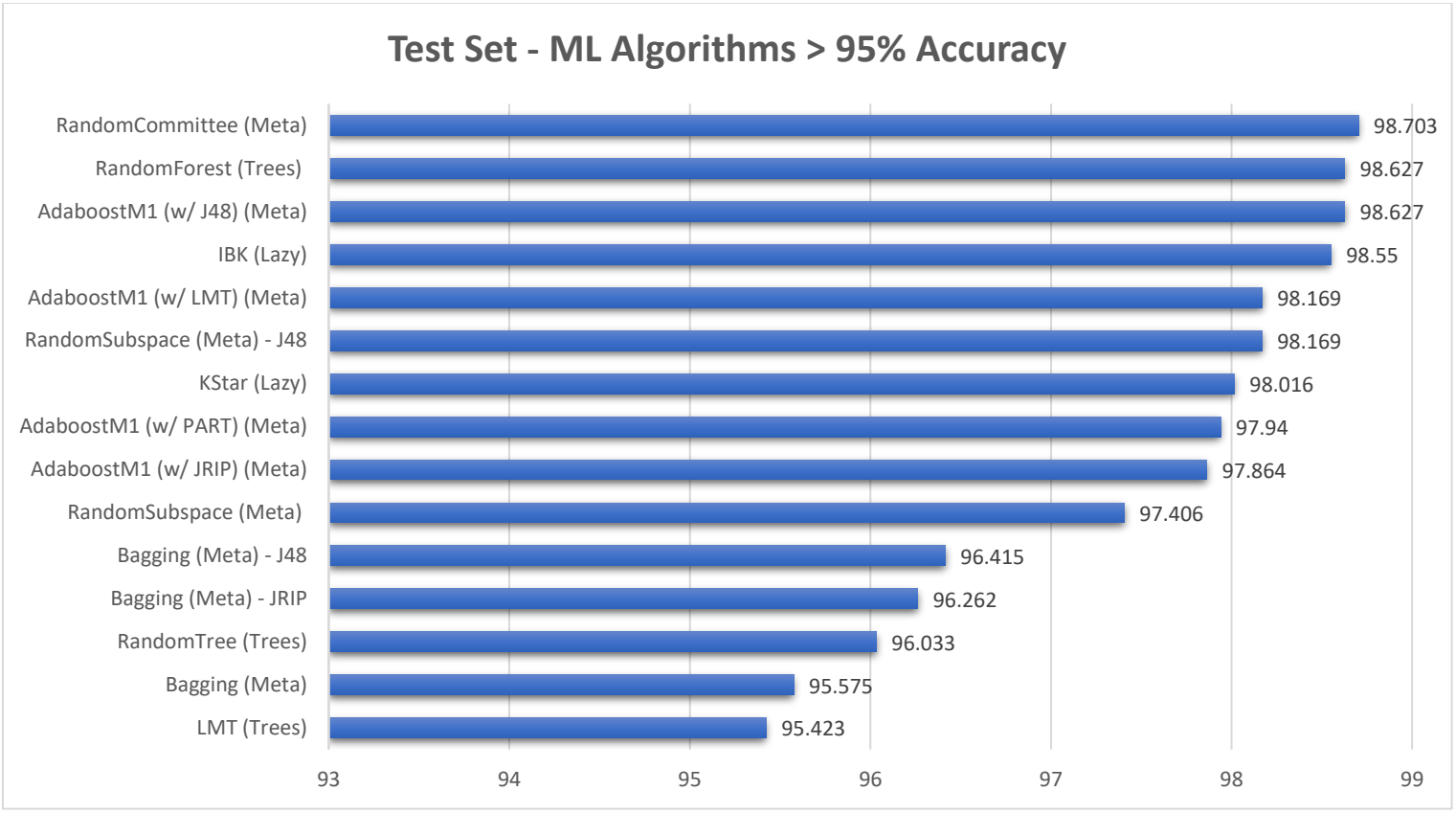


Figure 20 - Test Set - >95% Accuracy

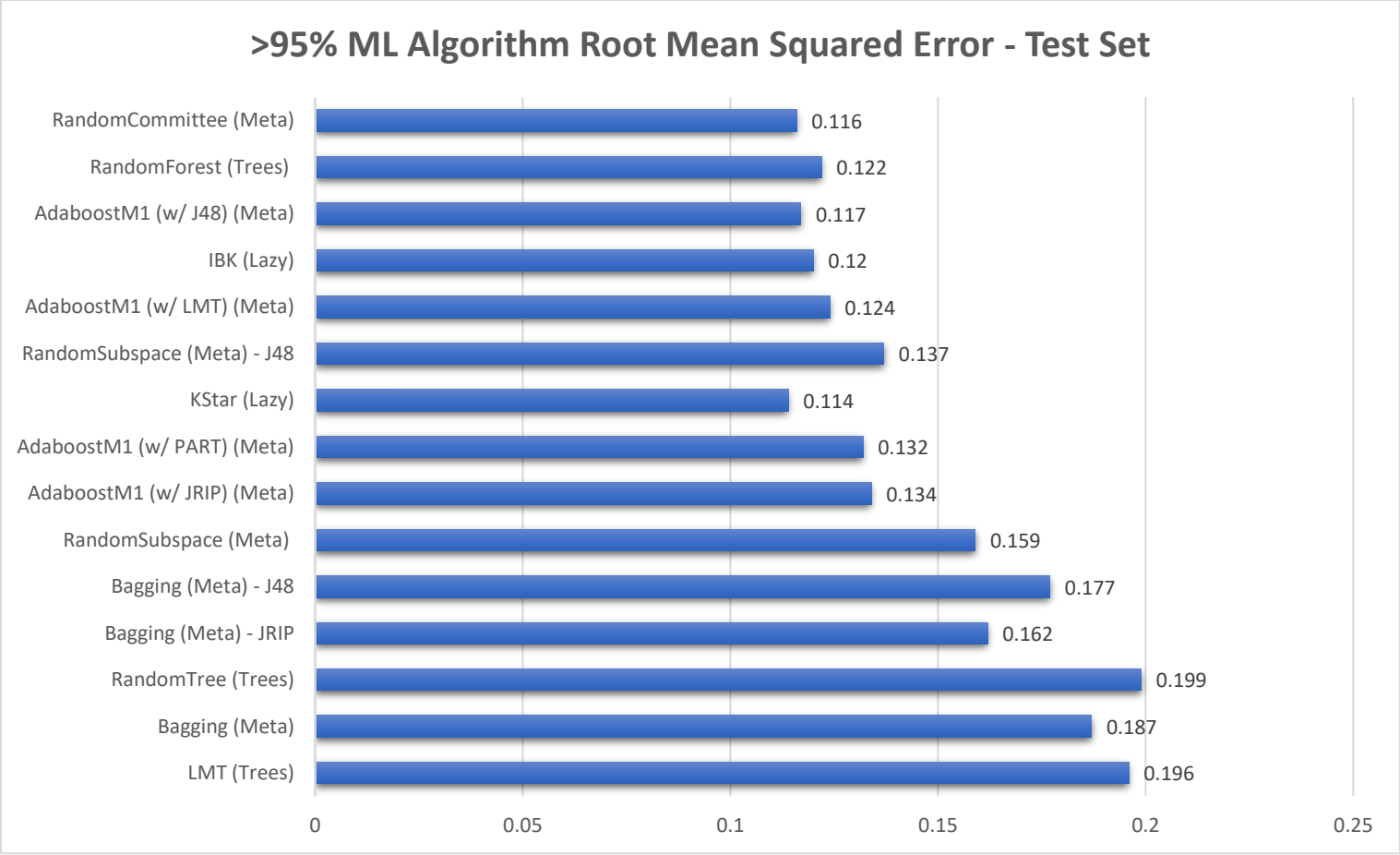


Figure 21 - Test Set - Root Mean Squared Error

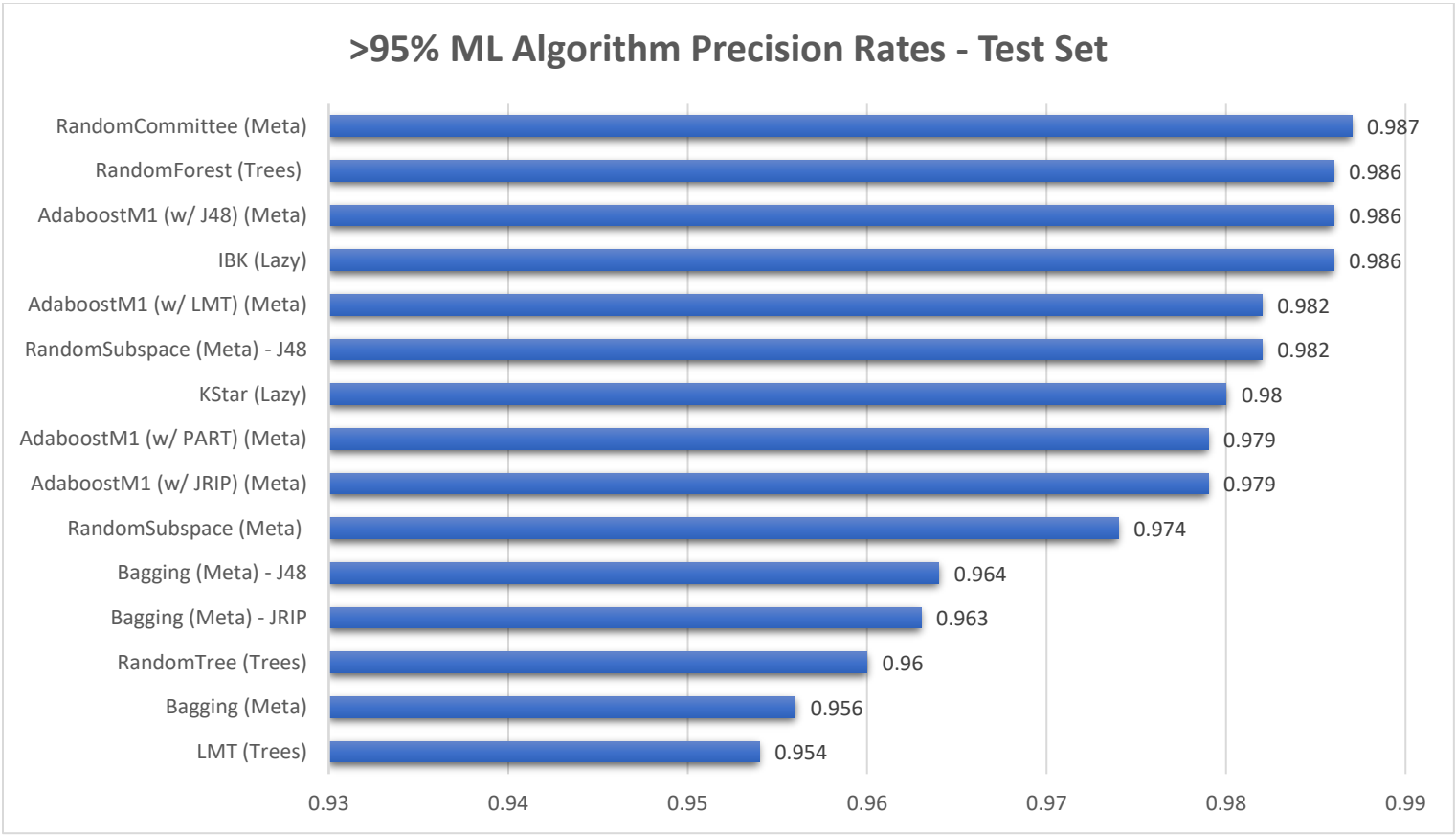


Figure 22 - Test Set - Precision Rates

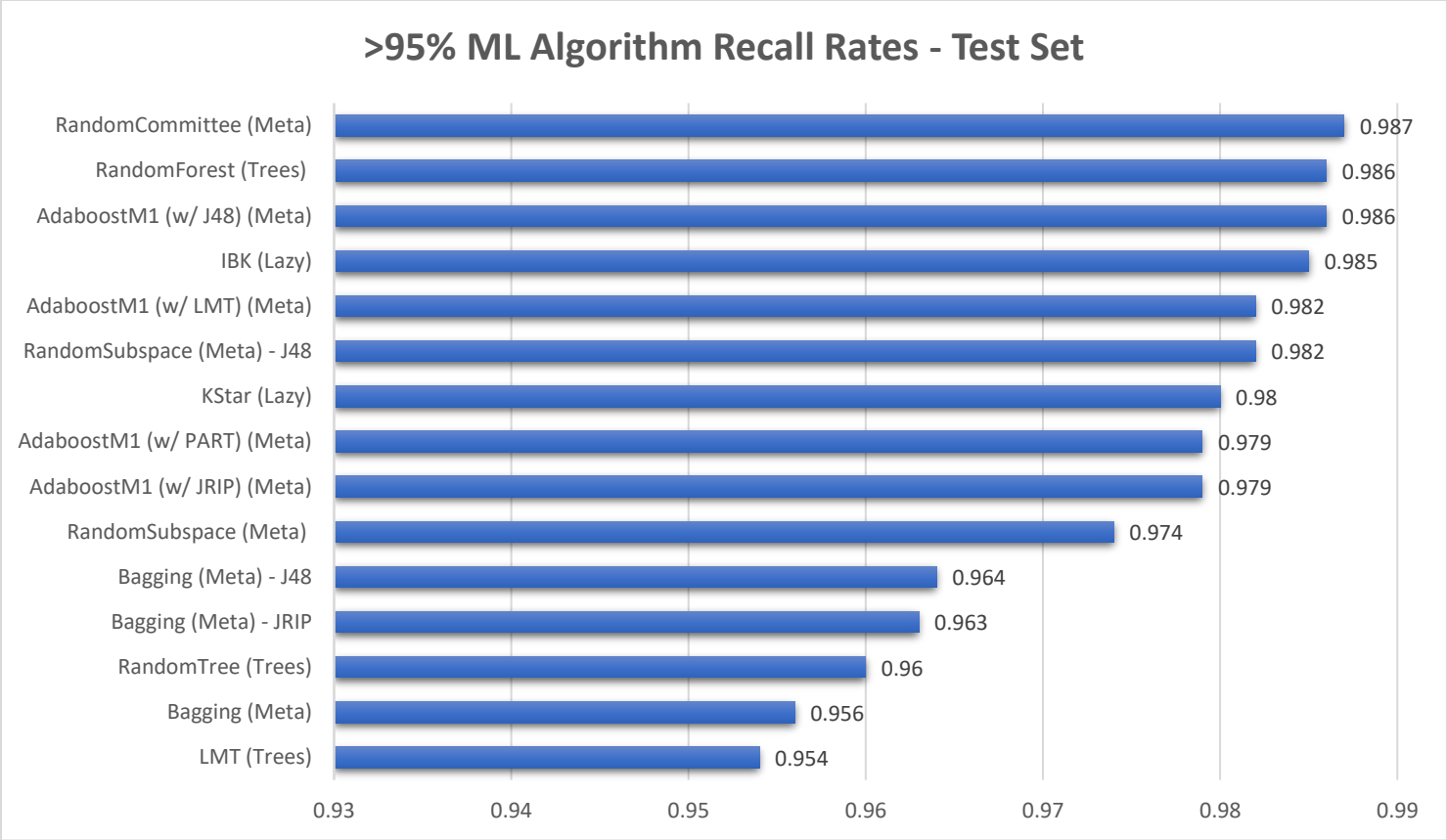


Figure 23 - Test Set - Recall Rates

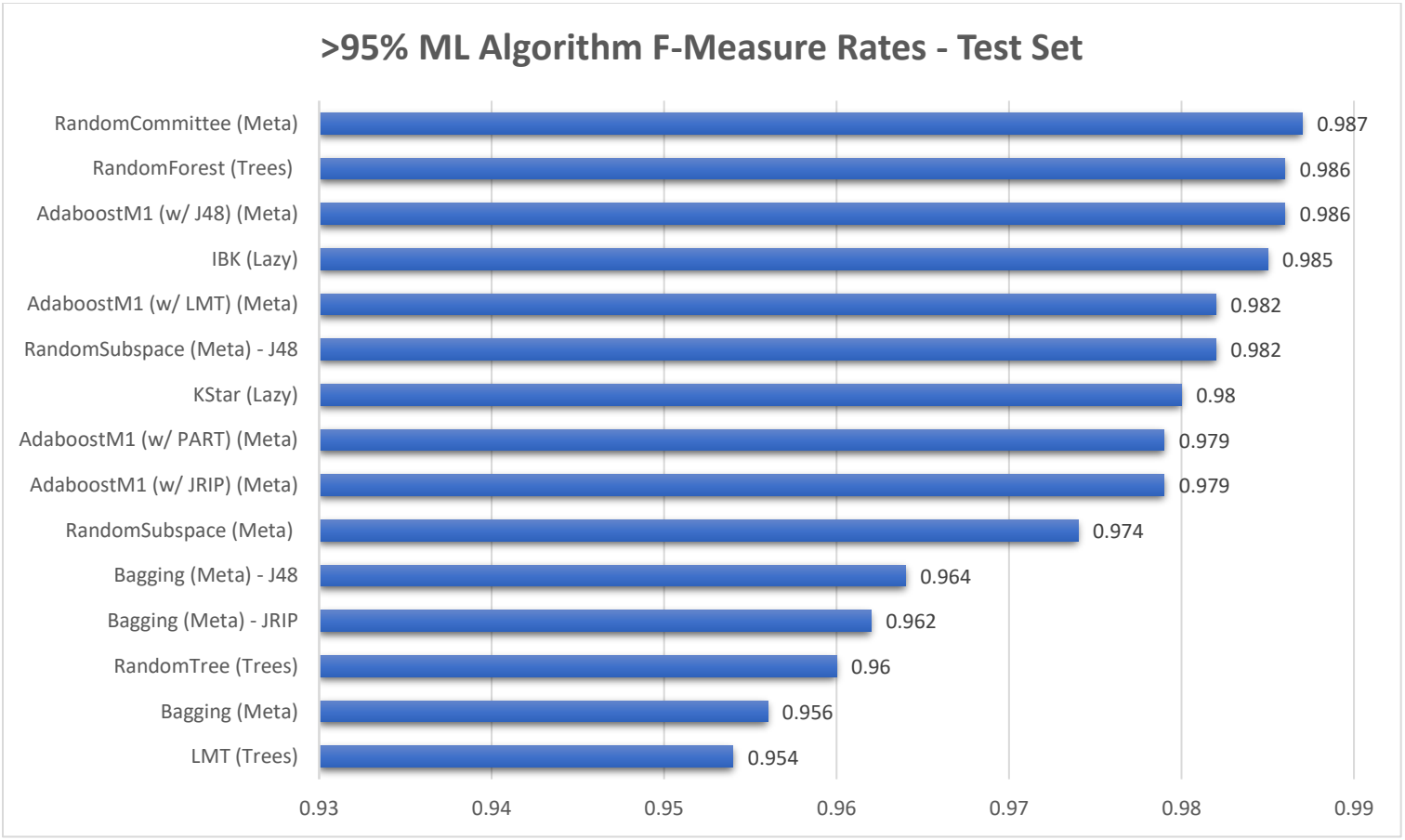


Figure 24 - Test Set - F-Measure Rates

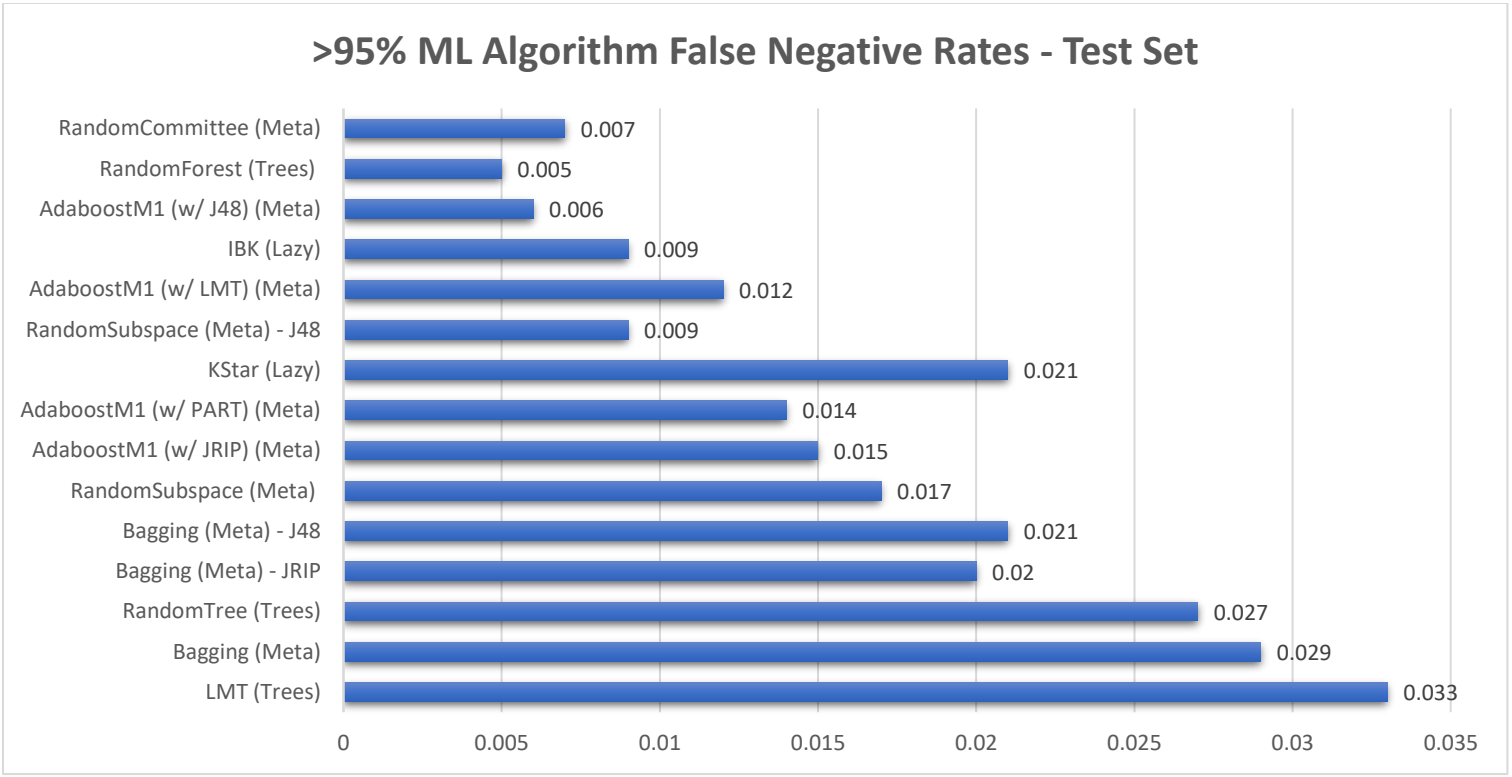


Figure 25 - Test Set - False Negative Rates

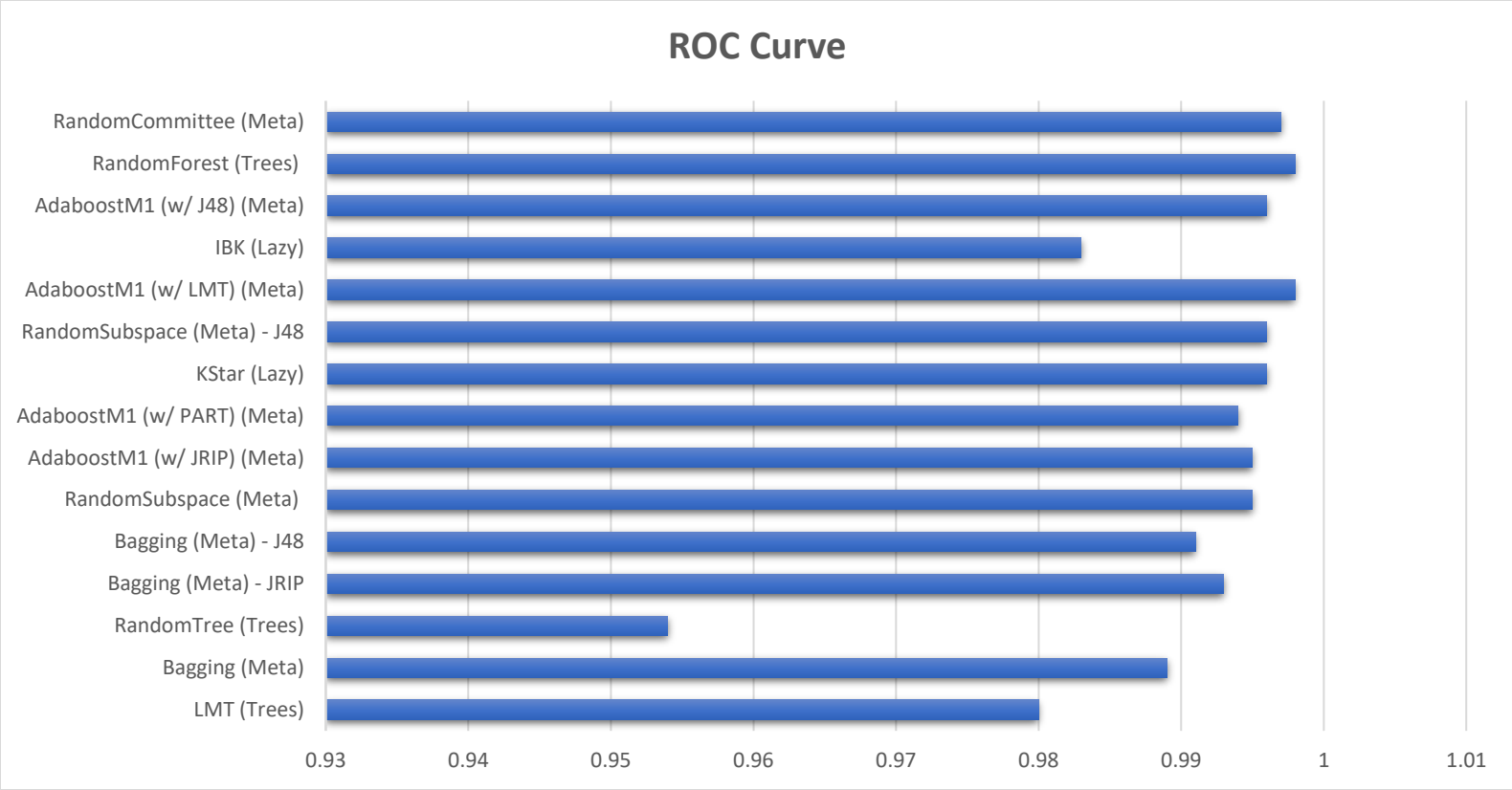


Figure 26 - ROC Curve

Given these metrics, the top three classifiers based purely on accuracy rates consisted of 2 meta classifiers and one tree classifier and had the following accuracy rates: RandomCommittee (98.703%), RandomForest (98.627%), and AdaboostM1+J48 (98.627%). The RandomCommittee classifier also outperformed RandomForest and AdaboostM1+J48 in regards to the highest Recall, Precision, and F-Measure, and the lowest Root Mean Squared Error. The RandomForest classifier outperformed RandomCommittee and AdaboosttM1+J48 in terms of the ROC Curve (.998 compared to .997 and .996 respectively) and also most importantly the FN rate (.005 compared to .007 and .006 respectively). Based on the catastrophic consequences of not identifying an attack quickly in an ICS environment, the FN rate should be given a great degree of weight when identifying the highest performing classifier. While the Time Taken (TT) rate is slightly higher for the RandomForest classifier compared to RandomCommittee and AdaboostM1+J48 (.08 seconds compared to .04 seconds and .07 seconds respectively), at least given the size of this dataset the increase in classification of attacks and minimization of FNs is worth the increase in computational time. However, with larger datasets or implementation in an IDS (and much larger volumes of data), this TT metric should be considered to a larger extent. Despite the relatively longer TT rate to evaluate the test data, the extremely high accuracy and the highest FN rate of the RandomForest algorithm suggests it is the highest performer. Also of note, of the 100 Random Trees generated in the RandomForest algorithm, the average size of the tree was 900.3 nodes (see Appendix J for model

output from the first Random Tree). Finally, conducting analysis through the attribute importance function within the RandomForest algorithm, it appears that 109 of the 128 attributes were used to classify an instance in at least one node of a Random Tree in the model (see Appendix J). This further confirms the results of the statistical methods, as classification is highly dependent on the utilization of a large number of variables.

CHAPTER 4: SUMMARY OF FINDINGS

This section presents a summary of the results and contributions of the research, issues and limitations experienced during the course of the research, and the potential for future work.

4.1 Results/Contributions

Given the current state of cybersecurity and its role in information operations and geopolitics, the research of exploitations of the electrical grid are extremely important and have the potential to directly affect national security. This work was focused on one possible exploitation that has been documented, but there are likely a myriad of other attack vectors as the attack surface is vast. Still, common TTPs such as spear phishing and an escalation of privileges through credential theft are being utilized to gain access to these networks, and training/education in proper cybersecurity practices within the ICS environment should be at the forefront of organizations in the energy sector. However, given that these TTPs are still effective and will likely remain effective, a hybrid IDS based off of synchrophasor data attack signatures is a necessity to quickly identify malicious activity and mitigate damaging effects. This research increased the knowledge base of remote tripping command injection attacks by demonstrating that multiple ML algorithms have a high degree of accuracy and are potentially good candidates to form the basis of a detection platform while statistical methods are

insufficient. The contribution to this specific area of cybersecurity/ICS is that this research suggests that detection axioms or rules cannot be formulated for remote tripping command injection attacks using the statistical methods of SLR and PCA, and thus should not be attempted when building signature based models of these attacks.

Also, given that different attacks could be characterized with optimal performance based on differing ML algorithms, this is a contribution to the body of knowledge regarding remote tripping command injection attacks. While the sheer volume of data from these measurement units is vast and likely unfathomable to a human operator, the application of ML methods can make sense of the data with a high degree of accuracy. This research broadened the scope of analysis by the application of numerous ML algorithms that had not been applied to analysis of remote tripping command injection attacks in previous work to the best knowledge of the author. Based on the preponderance of meta classifiers achieving 95% or greater accuracy on the test set, this research also suggests that more sophisticated ML algorithms have greater performance classification of remote tripping command injection attacks than simpler ML algorithms. This is related to the finding that the vast majority of variables were needed to formulate the RandomForest model which achieved the highest performance. Additionally, this research suggests different findings from Borges-Hink et al in which JRip+Adaboost was identified as the highest performing classifier for the entire dataset and the only classifier of seven that could classify at 95% or greater accuracy. For the subset explored in this work including only remote tripping command injection attacks and normal operational data, there are 15 algorithms that can accurately classify over a

95% rate. The work conducted by Borge-Hink et al had classification rates of approximately 95% and 79% over the entire binary dataset for JRipper+Adaboost and RandomForest respectively, while classification rates for remote tripping command injection attacks were 97.864% and 98.627% respectively. This suggests that specific attacks are classified with higher accuracy through an application of differing ML algorithms, and that in this case RandomForest is a better classifier for remote tripping command injection attacks than JRipper+Adaboost. To provide a synopsis of the results related to the Research Questions introduced in Section 1, each Research Question is addressed below.

4.1.1 Research Question 1 (RQ1) Results

The answer to RQ1 is that the statistical methods used in this research could not accurately model features of remote tripping command injection attacks on electrical grids due to a high degree of non-linearity and complexity in the dataset. It should be noted, that the original intent of this project was to determine simplistic detection axioms for remote tripping command injection attacks utilizing statistical methods. As described above, due to the composition of the dataset, these axioms were not possible to create given the statistical methods utilized. This is not to say that all statistical methods are incapable of producing axioms or rules that properly characterize data and thus provide an output as to whether at attack is occurring or not, but that the application of the methods SLR and PCA were unfruitful given the execution of scripts in R.

4.1.2 Research Question 2 (RQ2) Results

Through an execution of numerous ML algorithms in WEKA, it is clear that given the modified dataset that classification over 95% is possible for many algorithms (refer to Section 3.5). After splitting the dataset into training and test sets, accuracy metrics were determined via the use of the entire training set to build classifier models and evaluation through 10-fold Cross Validation. After execution of 10-fold Cross Validation was conducted an additional iteration of testing was conducted on an independent test set, it was determined that 15 algorithms achieved over 95% classification rates. Additionally, other metrics such as the FN rate suggest that the RandomForest model/detection axioms could be potentially be utilized in future work due to its high performance.

4.1.3 Research Question 3 (RQ3) Results

This is dependent on the individual IDS of an ICS environment, but this work has laid the groundwork for developing a signature based model for remote tripping command injection attacks based on detection axioms. Signature based models of

attacks have been utilized with high degrees of success and synchrophasor measurements based IDS have demonstrated a capacity to detect attacks that traditional IDS in ICS have not. The ML models formulated in this work are detection axioms or rules that characterize or describe remote tripping command injection attacks to a great degree of accuracy and could be utilized in future work with further validation.

4.1.4 Research Question 4 (RQ4) Results

GINA takes information from disparate domains (i.e., IT logs and OT sensor data), collates the data, and utilizes a unique form of modeling that has the ability to classify the data. While it is first necessary to construct a conceptual model based on detection axioms for implementation into GINA, there is a possibility given accuracy levels of the highest performing ML algorithms that the models/detection axioms established in this work could be incorporated in future work (see Section 4.3 for additional information about GINA) (Anderson, 2018).

4.2 Issues/Limitations

Given the omission of natural events and other attacks in the modified dataset used in this work, it is possible that the simplification of the dataset led to an increase in classification rates via the reduction in noise created by additional scenarios. The

modifications to the original dataset while extensive are still essentially subsetting the data into a more manageable and simpler dataset. While this was intentional to analyze only remote tripping command injection attacks in comparison to normal operations, the conditions are not representative of all activity/scenarios that could be experienced in an ICS environment. Also, in Borges-Hink et al. the datasets were randomly sampled at 1% of the original dataset to reduce the size and evaluate the effectiveness of small sample sizes. This must be noted in the final results, as the dataset used in this work was approximately 3x greater than this dataset. Additionally, infinite values were modified in the dataset to facilitate analysis in R as detailed in Section 3.4 and were kept for the purpose of analysis comparison and consistency across ML and statistical methods.

Another associated issue in both the original dataset and the modified dataset used in this research is class imbalance. As previously mentioned in Section 3.4, there were 8,737 attacks and 4,405 normal operational instances. This does not likely mirror the daily operational data distribution observed in an ICS environment due to the large proportion of attacks compared to normal operations. Therefore, if a more realistic dataset was evaluated with the top performing ML algorithms identified in this work, the results could differ significantly. In future work, methods of reducing class imbalance could be employed on this dataset and the resulting ML algorithm accuracy metrics could be comparatively evaluated (Caulkins, 2018; Lathrop, 2018; Wiegand, 2018).

A peripheral technical issue associated with this work is that synchrophasor units can be targeted by attackers, as IECC C37.1118 (the protocol for synchrophasor data

communication) does not support any authentication and thus readings of sensors could be manipulated and render a system obsolete through exploitation of Man-in-the-Middle types of attacks (Borges-Hink et al., 2014; Yang, McLaughlin, Sezer, Littler, Pranggono, Brogan, and Wong, 2013). This was outside the scope of this research, which sought to utilize synchrophasor data to detect remote tripping command injection attacks but did not address exploitation of the PMUs themselves. Another technical issue is that this dataset reflects a non-pilot directional over current relay protection scheme in multiple source circuits (Pan et al., March 2015). While this type of protection was utilized in the original study due that makes up the majority of the electric transmission system, other circuits exist such as loop and radial circuits are also present in the electrical grid and the results/models obtained in this work could be nonapplicable to systems with these components (Pan et al., March 2015).

There are also potential WEKA comparison issues with the previous work conducted by Borges-Hink et al. After consulting with the lead author about parameters and reading datatypes in to WEKA, due to a loss of the original results an approximation of accuracy, precision, recall, and F-Measure was used based off interpretation of graphs in the article. Additionally, the author stated that he did not believe he used IT component logs for his paper, which would indicate IT variables have little to no effect on classification accuracy in the original work. While this would be a unique contribution to the body of research for this work, the original work does directly refute this by including the IT logs as part of Results Section C (Borges-Hink et al., 2014). Finally, to compare training to test sets an InputMappedClassifier application was used in WEKA

so that differing value ranges for R1-R4.PA.Z could be utilized during the final phase of testing. The original author stated that he could not recall how he had compared training to test sets in WEKA and could not confirm or deny he used the InputMappedClassifier (Borges-Hink, 2018). Also, regarding the WEKA evaluation methodology, the ML algorithms are stochastic methods that were only run once during analysis in this work (Wiegand, 2018). It is likely that reevaluating the ML algorithms with differing Random Seed values would result in differing metrics and could potentially give a more complete picture of a model's ability to classify remote tripping command injection attacks (Wiegand, 2018).

Another associated issues about the PMU data in IDS is consternation regarding data storage. If this small testbed architecture is generating over 13,000 datapoints in a short amount of time, there are undoubtedly related issues regarding storage of the data, policies related as to how long to keep the data, and what criteria would need to be met to keep data long term at the present time (Redwood, 2018). This is likely a reason for energy companies to rely on more traditional IDS without synchrophasor measurements incorporated, but with the increase in computational power and storage capacity there is a possibility that IDS incorporating synchrophasor data will become more widespread (Redwood, 2018).

4.3 Future Work

While this research suggests that ML models/axioms were successful in identifying remote tripping command injection attacks, further research and additional testing should be done to confirm these results. Testing actual remote tripping command injections outside of a laboratory or testbed environment is likely not possible due to the ramifications of the attack, but the ML algorithms/detection axioms could be validated on multiple independent datasets. A future expansion of this work could be to incorporate the other attack and natural scenarios in a binary dataset and observe if the accuracy metrics were replicated. Further parameter calibration of the highest performing ML algorithms identified could be another potential direction for expanding this work. An analysis given these axioms/ML models and a different testbed architecture could be utilized to find if accuracy metrics are confirmed in a completely independent dataset (i.e., one that is not a subset of existing data). If the detection axioms/attack models are confirmed in separate independent testing, this increases the viability of integrating this research into a hybrid IDS that incorporates synchrophasor data.

There is also a potential for Vector Relational Data Modeling (VRDM) to be utilized in future work as it has been shown to be effective in real time multivariate analysis (Dougherty, 2017). VRDM is the underlying framework or the programming language engine powering a software solution called the Global Information Network Architecture (GINA), which has been utilized for such applications as identifying and comparing Naval Energy Weapons systems, enabling system interoperability in smart mobile system services of network decision support systems, geospatial mapping of IP

addresses, and automated threat analysis via IDS (Dougherty, 2017; Dolk, Busalacchi, Anderson and Tinsley, 2012, ,and Cohort 19, Team Bravo, 2013). GINA recognizes relationships between data objects specified by the user and does not require coding, which offers benefits in time conservation and also decreases errors associated with physically coding the model and encourages a focus on the design of overarching attributes and functionality of the model, as opposed to hard coded mechanics (Cohort 19, Team Bravo, 2013; Dolk et al., 2012). Additionally, a model can be easily recalibrated or modified and does not force a complete recoding effort (Cohort 19, Team Bravo, 2013; Dolk et al., 2012).

As demonstrated in the previous sections, ICS environments have many complex relationships that increases the difficulty of constructing an accurate model. GINA provides the user with an interface to model these complex interactions across domains that is both intuitive and not resource intensive, which is especially appropriate given the nature and interconnected relationships of vast amounts of IT and OT components in ICS. Conceptual relationship models can be easily implemented given a user's knowledge of a domain and ability to discern interactions between components. While it was not possible to integrate the ML models/detection axioms established in this research due to time constraints stemming from complications during dataset analysis, the ML models/detection axioms established in this work could potentially be implemented in GINA in the future for further analysis of remote tripping command injection attacks (Anderson, 2018).

APPENDIX A. PRINCIPAL COMPONENT ANALYSIS R SCRIPTS

The scripts below in R detail the process of executing Principal Component Analysis given the standardized dataset (Coghlan, 2013 and Wiegand, 2018). Outputs include the screeplot, standard deviation/proportion of variance, and coefficient analysis of the 25 rotations.

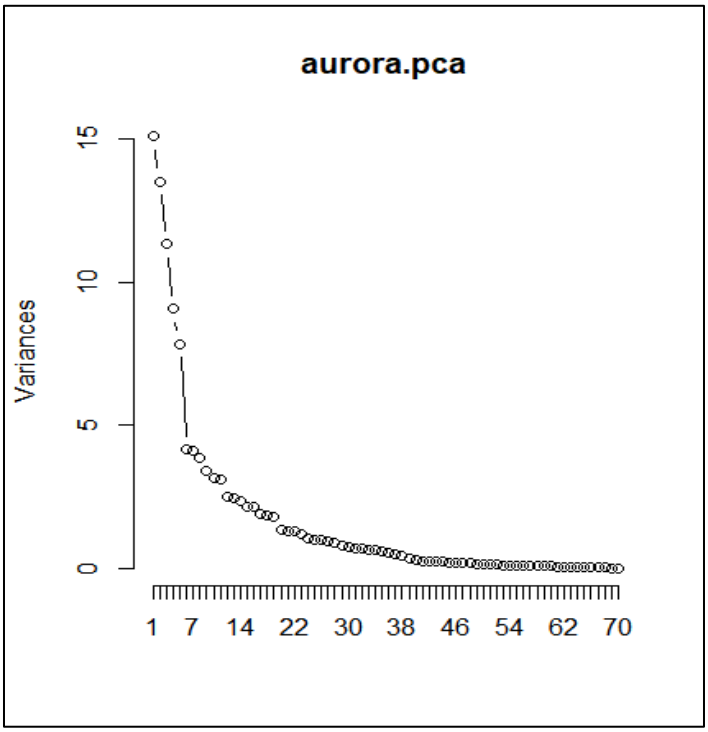
```
#Standardize in order to compare variables which have vastly different variances; should only apply to non-factor variables (ie, no logs or marker);
#but how does this affect the overall dataset?  Should I standardize only some of the values, or the whole set?
#We want output for mean to be close to 0, and standard deviations to be 1

standardizeAll <- as.data.frame(scale(aurora[1:116])) ## Only standardized through the numeric data (column 116)...can't with nominal (logs and marker)

sapply(standardizeAll,mean)

sapply(standardizeAll,sd)

aurora.pca <- prcomp(standardizeAll)

summary(aurora.pca) #outputs sdev, proporation of variance, and cumulative proportion, can also do commands below individually
###for this dataset, PC1 is .1303 for cumulative proportion...to get to over 95%, we need to go to PC34, to over 99% PC51 (actually
## PC52, PC51 is 99.0000%)
## This is an alternate method to scree decide how many we should retain is based off the minimum amount of variance;
##Kaiser's Criterion - another way to decide on which components to retain
## We should only retain components whose variance is above 1 (when pca applied to standardized data); this means through PC25 - GO WITH THAT!

qplot(aurora.pca, x=PC1, y=PC2)

autoplot(prcomp(df), data=aurora.pca, color='marker')

aurora.pca$sdev #individual command for sdev

sum((aurora.pca$sdev)^2) #sum of the variances gives us the total variance (in this case, 116 since I only had variables in columns 1-116)


## Deciding How Many to Retain - Scree Plot - change in slope will probably dicatate which components should be retaind
## usually highest value of variance should be retained; based on this example, Columns 1-6, as the slope levels off after

##screeplot(aurora.pca, type="lines")  ##original command from tutorial, use the one below for this graph

screeplot(aurora.pca, npcs=70, type="lines") #modified number of components (NPCS) plotted
## Slope levels off around 24-25
```

```
##Obtaining the loadings - first column indicates 1st PC, second is 2nd, etc
## NEED TO CHECK ALL THE PCS THROUGH 25!!!  SEE IF ALL THE VARIABLES ARE PRESENT IN THE LOADINGS - IF SO, THAT INDICATES THAT THEY
## ALL NEED TO BE INCLUDED

aurora.pca$rotation[,1]
##OUTPUT: PC1 = -2.34xe-01*Z1(Standardized R1.PA1.VH)...
aurora.pca$rotation[,2]
aurora.pca$rotation[,3]
aurora.pca$rotation[,4]
aurora.pca$rotation[,5]
aurora.pca$rotation[,6]
aurora.pca$rotation[,7]
aurora.pca$rotation[,8]
aurora.pca$rotation[,9]
aurora.pca$rotation[,10]
aurora.pca$rotation[,11]
aurora.pca$rotation[,12]
aurora.pca$rotation[,13]
aurora.pca$rotation[,14]
aurora.pca$rotation[,15]
aurora.pca$rotation[,16]
aurora.pca$rotation[,17]
aurora.pca$rotation[,18]
aurora.pca$rotation[,19]
aurora.pca$rotation[,20]
aurora.pca$rotation[,21]
aurora.pca$rotation[,22]
aurora.pca$rotation[,23]
aurora.pca$rotation[,24]
aurora.pca$rotation[,25]

##Square of the loadings sum to 1, which is constraint used in calculating loadings
sum(aurora.pca$rotation[,1]^2)
```

```r
#Calculating values of PC1 for each sample (row in the dataset)
calcpc <- function(variables,loadings)
{
  # find the number of samples in the data set
  as.data.frame(variables)
  numsamples <- nrow(variables)
  # make a vector to store the component
  pc <- numeric(numsamples)
  # find the number of variables
  numvariables <- length(variables)
  # calculate the value of the component for each sample
  for (i in 1:numsamples)
  {
    valuei <- 0
    for (j in 1:numvariables)
    {
      valueij <- variables[i,j]
      loadingj <- loadings[j]
      valuei <- valuei + (valueij * loadingj)
    }
    pc[i] <- valuei
  }
  return(pc)
}

calcpc(standardizeAll, aurora.pca$rotation[,1])

##Output of values for all data entries (over 13000)...for PC1?

##check that the values match the principal components are same as we exectued earlier with prcomp
aurora.pca$x[,1]
```

```r
### - DR WIEGAND CODE - 7MAY18

####### - UPDATED WITH ERRORS LINES TAKEN OUT, AND VARS CHANGED FROM 91 to 116

aurora.pca$rotation[1,1:25]

head(as.data.frame(aurora.pca$rotation[1:116,1:25])
    )

boxplot(aurora.pca$rotation[1,1:25])
boxplot(aurora.pca$rotation[2,1:25])
boxplot(aurora.pca$rotation[3,1:25])
boxplot(aurora.pca$rotation[116,1:25])
as.vector(aurora.pca$rotation[1:116,1:25])

boxplot(as.vector(aurora.pca$rotation[1:116,1:25]), notch=T)

## t.test(as.vector(aurora.pca$rotation[1:91,1:25])$conf.int
## ERROR

t.test(as.vector(aurora.pca$rotation[1:116,1:25]))$conf.int
## OUTPUT: .95

IQR(as.vector(aurora.pca$rotation[1:116,1:25]))
## OUTPUT: 0.05212563
## ORIGINAL OUTPUT: 0.05475931
## LOWER THAN FIRST ITERATION WITH DR W

 summary(as.vector(aurora.pca$rotation[1:116,1:25]))
## OUTPUT: Min.    1st Qu.     Median       Mean   3rd Qu.      Max.
##      -0.599949 -0.027454 -0.001526 -0.004642  0.024671  0.720123
## ORIGINAL OUTPUT: Min.    1st Qu.     Median       Mean   3rd Qu.      Max.
##                -0.599949 -0.030110 -0.001917 -0.006532  0.024650  0.720123
## MIN & MAX THE SAME, ALL OTHERS slightly less other than 3rd quartile

x = aurora.pca$rotation[1:116,1:25]
```

```
dim(x)

x = abs(aurora.pca$rotation[1:116,1:25])

greaterThan <- function(val) { return (val > 0.15)+0; }
greaterThan(2)
## OUTPUT: TRUE
## ORIGINAL OUTPUT: TRUE

greaterThan <- function(val) { return ((val > 0.15)+0); }
greaterThan(2)
## OUTPUT: 1
## ORIGINAL OUTPUT: 1

greaterThan(0.01)
## OUTPUT: 0
## ORIGINAL OUTPUT: 0

apply(x, 1, greaterThan)
## OUTPUT APPEARS TO BE SAME AS ORIGINAL...

dim(apply(x,1,greaterThan))
## OUTPUT: 25 116 (reflects all variables...orig output had 91)

x > 0.15
## OUTPUT APPEARS TO BE SAME AS ORIG...

(x > 0.15)+0
## OUTPUT APPEARS SAME AS ORIG

apply((x > 0.15)+0, 2, sum)
## OUTPUT:
##  PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10 PC11 PC12 PC13 PC14 PC15 PC16 PC17 PC18 PC19 PC20 PC21 PC22 PC23 PC24 PC25
##   16   13   16   14   14   12    4   17   14   11   10   10    6   12   14    8   10    8    9   12    6   11   10    8    6
## ORGINAL OUTPUT:
##  PC1  PC2  PC3  PC4  PC5  PC6  PC7  PC8  PC9 PC10 PC11 PC12 PC13 PC14 PC15 PC16 PC17 PC18 PC19 PC20 PC21 PC22 PC23 PC24 PC25
##   13    7   16   11   12   11    3   16    8    9    8    5    5   10   11    7    9    7    8   10    3    8    6    6    3
```



```
> aurora.pca <- prcomp(standardizeAll)
                          PC1    PC2    PC3     PC4     PC5     PC6     PC7
Standard deviation     3.884 3.6745 3.3648 3.00857 2.79950 2.03728 2.02047
Proportion of Variance 0.130 0.1164 0.0976 0.07803 0.06756 0.03578 0.03519
Cumulative Proportion  0.130 0.2464 0.3440 0.42206 0.48962 0.52540 0.56060
                          PC8     PC9   PC10    PC11    PC12    PC13    PC14
Standard deviation     1.96414 1.85016 1.7794 1.76289 1.57755 1.56727 1.53027
Proportion of Variance 0.03326 0.02951 0.0273 0.02679 0.02145 0.02118 0.02019
Cumulative Proportion  0.59385 0.62336 0.6507 0.67745 0.69890 0.72008 0.74027
                         PC15    PC16    PC17   PC18    PC19   PC20   PC21
Standard deviation     1.47435 1.46207 1.37590 1.3582 1.34295 1.1652 1.1399
Proportion of Variance 0.01874 0.01843 0.01632 0.0159 0.01555 0.0117 0.0112
Cumulative Proportion  0.75901 0.77743 0.79375 0.8097 0.82520 0.8369 0.8481
                         PC22    PC23    PC24    PC25    PC26    PC27    PC28
Standard deviation     1.13555 1.08498 1.03020 1.00538 0.99020 0.98280 0.94094
Proportion of Variance 0.01112 0.01015 0.00915 0.00871 0.00845 0.00833 0.00763
Cumulative Proportion  0.85923 0.86937 0.87852 0.88724 0.89569 0.90402 0.91165
```

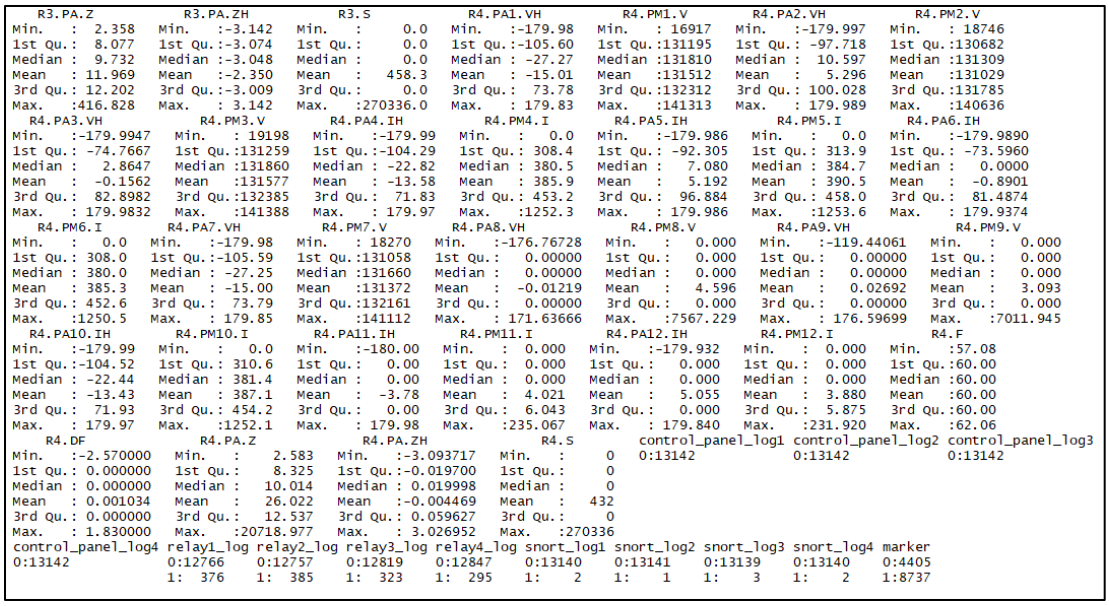
```
> x > 0.15
            PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8   PC9  PC10  PC11  PC12  PC13  PC14  PC15  PC16  PC17  PC18  PC19  PC20  PC21  PC22  PC23  PC24  PC25
R1.PA1.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM1.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA2.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM2.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA3.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM3.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA4.IH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM4.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA5.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM5.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA6.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM6.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA7.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM7.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA8.VH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
R1.PM8.V   FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA9.VH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE  TRUE
R1.PM9.V   FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE  TRUE FALSE FALSE
R1.PA10.IH  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PM10.I  FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA11.IH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
R1.PM11.I  FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.PA12.IH FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
R1.PM12.I  FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.F       FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R1.DF      FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
R1.PA.Z    FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE  TRUE FALSE
R1.PA.ZH   FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE
R1.S       FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA1.VH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM1.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA2.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM2.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA3.VH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM3.V   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA4.IH   TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PM4.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA5.IH  FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
R2.PM5.I   FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
R2.PA6.IH  FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
 [ reached getOption("max.print") -- omitted 76 rows ]
```

APPENDIX B. INITIAL LOADING AND DATASET ANALYSIS

The loading, initial summary, and structure of the data was found through the execution of the R Scripts below (Boone, 2010, Stack Overflow Thread, 2014, Stack Overflow Thread, 2015, Mollie, 2013, and Wiegand, 2018). Outputs for the data summary and structure are shown below the R scripts.

```
aurora <- read.csv(
  file.choose(),
  header = TRUE, sep = ',',
  colClasses=c("marker"="factor", "control_panel_log1"="factor", "control_panel_log2"="factor",
               "control_panel_log3"="factor", "control_panel_log4"="factor", "relay1_log"="factor",
               "relay2_log"="factor", "relay3_log"="factor",
               "relay4_log"="factor", "snort_log1"="factor", "snort_log2"="factor", "snort_log3"="factor", "snort_log4"="factor")
)


### Check data to ensure it's been loaded
summary(aurora)

### Check datatypes - From - https://stackoverflow.com/questions/21125222/determine-the-data-types-of-a-data-frames-columns
## From - https://stackoverflow.com/questions/28895044/list-output-truncated-how-to-expand-listed-variables-with-str-in-r
str(aurora, list.len=129)
```

```
   R2.PM7.V         R2.PA8.VH            R2.PM8.V          R2.PA9.VH            R2.PM9.V          R2.PA10.IH         R2.PM10.I
Min.   : 95429   Min.   :-163.97479   Min.   :   0.000   Min.   :-130.26196   Min.   :   0.0000   Min.   :-180.00   Min.   :   0.0
1st Qu.:129002   1st Qu.:   0.00000   1st Qu.:   0.000   1st Qu.:   0.00000   1st Qu.:   0.0000   1st Qu.: -65.26   1st Qu.: 315.6
Median :130173   Median :   0.00000   Median :   0.000   Median :   0.00000   Median :   0.0000   Median :  18.79   Median : 387.9
Mean   :129133   Mean   :  -0.00951   Mean   :   4.361   Mean   :  -0.01115   Mean   :   0.9882   Mean   :  15.03   Mean   : 391.3
3rd Qu.:131082   3rd Qu.:   0.00000   3rd Qu.:   0.000   3rd Qu.:   0.00000   3rd Qu.:   0.0000   3rd Qu.:  98.36   3rd Qu.: 461.6
Max.   :138161   Max.   : 167.69902   Max.   :7031.533   Max.   : 125.81269   Max.   :2030.9349   Max.   : 180.00   Max.   :1267.9
  R2.PA11.IH        R2.PM11.I         R2.PA12.IH         R2.PM12.I           R2.F           R2.DF             R2.PA.Z
Min.   :-179.893   Min.   :  0.000   Min.   :-179.986   Min.   :  0.000   Min.   :58.04   Min.   :-1.5100000   Min.   :  2.365
1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:60.00   1st Qu.: 0.0000000   1st Qu.:  8.072
Median :   0.000   Median :  0.000   Median :   0.000   Median :  0.000   Median :60.00   Median : 0.0000000   Median :  9.743
Mean   :   2.744   Mean   :  4.154   Mean   :  -4.665   Mean   :  4.047   Mean   :60.00   Mean   : 0.0006955   Mean   : 13.996
3rd Qu.:   0.000   3rd Qu.:  6.043   3rd Qu.:   0.000   3rd Qu.:  6.043   3rd Qu.:60.00   3rd Qu.: 0.0000000   3rd Qu.: 12.216
Max.   : 179.909   Max.   :217.627   Max.   : 179.953   Max.   :223.091   Max.   :64.31   Max.   : 1.5300000   Max.   :797.606
   R2.PA.ZH          R2.S          R3.PA1.VH          R3.PM1.V         R3.PA2.VH          R3.PM2.V         R3.PA3.VH
Min.   :-3.142   Min.   :     0   Min.   :-179.97   Min.   : 95404   Min.   :-179.989   Min.   : 95379   Min.   :-179.9890
1st Qu.:-3.075   1st Qu.:     0   1st Qu.:-108.04   1st Qu.:128726   1st Qu.: -95.243   1st Qu.:128676   1st Qu.: -73.3429
Median :-3.049   Median :     0   Median : -29.13   Median :129829   Median :  15.023   Median :129804   Median :   1.7647
Mean   :-2.297   Mean   :   432   Mean   : -15.01   Mean   :128818   Mean   :   5.113   Mean   :128788   Mean   :   0.9927
3rd Qu.:-3.009   3rd Qu.:     0   3rd Qu.:  73.43   3rd Qu.:130732   3rd Qu.:  97.709   3rd Qu.:130707   3rd Qu.:  81.0707
Max.   : 3.142   Max.   :270336   Max.   : 179.99   Max.   :137702   Max.   : 179.989   Max.   :137677   Max.   : 179.9947
   R3.PM3.V         R3.PA4.IH         R3.PM4.I         R3.PA5.IH          R3.PM5.I         R3.PA6.IH          R3.PM6.I
Min.   : 95454   Min.   :-179.99   Min.   :   0.0   Min.   :-179.989   Min.   :   0.0   Min.   :-179.9947   Min.   :   0.0
1st Qu.:128751   1st Qu.: -65.50   1st Qu.: 312.6   1st Qu.: -90.613   1st Qu.: 318.2   1st Qu.: -98.3482   1st Qu.: 311.5
Median :129880   Median :  18.68   Median : 385.4   Median :  -7.151   Median : 389.8   Median :   0.0000   Median : 384.7
Mean   :128856   Mean   :  15.00   Mean   : 390.4   Mean   :  -5.403   Mean   : 394.9   Mean   :   0.2343   Mean   : 389.7
3rd Qu.:130757   3rd Qu.:  99.21   3rd Qu.: 458.3   3rd Qu.:  69.809   3rd Qu.: 462.5   3rd Qu.: 103.0751   3rd Qu.: 457.6
Max.   :137727   Max.   : 179.99   Max.   :1264.6   Max.   : 179.983   Max.   :1266.0   Max.   : 179.9890   Max.   :1264.0
   R3.PA7.VH         R3.PM7.V         R3.PA8.VH          R3.PM8.V         R3.PA9.VH          R3.PM9.V         R3.PA10.IH
Min.   :-179.99   Min.   : 95404   Min.   :-164.45035   Min.   :   0.000   Min.   :-129.95829   Min.   :   0.0000   Min.   :-179.99
1st Qu.:-108.16   1st Qu.:128726   1st Qu.:   0.00000   1st Qu.:   0.000   1st Qu.:   0.00000   1st Qu.:   0.0000   1st Qu.: -65.59
Median : -29.15   Median :129854   Median :   0.00000   Median :   0.000   Median :   0.00000   Median :   0.0000   Median :  18.45
Mean   : -15.09   Mean   :128820   Mean   :  -0.00972   Mean   :   4.325   Mean   :  -0.01079   Mean   :   0.9902   Mean   :  14.85
3rd Qu.:  73.38   3rd Qu.:130732   3rd Qu.:   0.00000   3rd Qu.:   0.000   3rd Qu.:   0.00000   3rd Qu.:   0.0000   3rd Qu.:  99.36
Max.   : 179.97   Max.   :137702   Max.   : 167.45837   Max.   :6744.710   Max.   : 128.65767   Max.   :2056.0081   Max.   : 179.98
  R3.PM10.I         R3.PA11.IH         R3.PM11.I         R3.PA12.IH         R3.PM12.I          R3.F             R3.DF
Min.   :   0.0   Min.   :-179.995   Min.   :  0.000   Min.   :-179.914   Min.   :  0.000   Min.   :58.04   Min.   :-1.5100000
1st Qu.: 314.6   1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:60.00   1st Qu.: 0.0000000
Median : 386.4   Median :   0.000   Median :  0.000   Median :   0.000   Median :  0.000   Median :60.00   Median : 0.0000000
Mean   : 391.5   Mean   :   2.966   Mean   :  4.091   Mean   :  -4.256   Mean   :  3.931   Mean   :60.00   Mean   : 0.0006559
3rd Qu.: 459.1   3rd Qu.:   0.000   3rd Qu.:  6.043   3rd Qu.:   0.000   3rd Qu.:  5.860   3rd Qu.:60.00   3rd Qu.: 0.0000000
Max.   :1264.9   Max.   : 179.892   Max.   :239.508   Max.   : 179.989   Max.   :230.902   Max.   :64.31   Max.   : 1.5300000
```



```
    R3.PA.Z           R3.PA.ZH           R3.S           R4.PA1.VH          R4.PM1.V        R4.PA2.VH          R4.PM2.V
Min.   :  2.358   Min.   :-3.142   Min.   :     0.0   Min.   :-179.98   Min.   : 16917   Min.   :-179.997   Min.   : 18746
1st Qu.:  8.077   1st Qu.:-3.074   1st Qu.:     0.0   1st Qu.:-105.60   1st Qu.:131195   1st Qu.: -97.718   1st Qu.:130682
Median :  9.732   Median :-3.048   Median :     0.0   Median : -27.27   Median :131810   Median :  10.597   Median :131309
Mean   : 11.969   Mean   :-2.350   Mean   :   458.3   Mean   : -15.01   Mean   :131512   Mean   :   5.296   Mean   :131029
3rd Qu.: 12.202   3rd Qu.:-3.009   3rd Qu.:     0.0   3rd Qu.:  73.78   3rd Qu.:132312   3rd Qu.: 100.028   3rd Qu.:131785
Max.   :416.828   Max.   : 3.142   Max.   :270336.0   Max.   : 179.83   Max.   :141313   Max.   : 179.989   Max.   :140636
   R4.PA3.VH          R4.PM3.V        R4.PA4.IH          R4.PM4.I        R4.PA5.IH          R4.PM5.I        R4.PA6.IH
Min.   :-179.9947   Min.   : 19198   Min.   :-179.99   Min.   :   0.0   Min.   :-179.986   Min.   :   0.0   Min.   :-179.9890
1st Qu.: -74.7667   1st Qu.:131259   1st Qu.:-104.29   1st Qu.: 308.4   1st Qu.: -92.305   1st Qu.: 313.9   1st Qu.: -73.5960
Median :   2.8647   Median :131860   Median : -22.82   Median : 380.5   Median :   7.080   Median : 384.7   Median :   0.0000
Mean   :  -0.1562   Mean   :131577   Mean   : -13.58   Mean   : 385.9   Mean   :   5.192   Mean   : 390.5   Mean   :  -0.8901
3rd Qu.:  82.8982   3rd Qu.:132385   3rd Qu.:  71.83   3rd Qu.: 453.2   3rd Qu.:  96.884   3rd Qu.: 458.0   3rd Qu.:  81.4874
Max.   : 179.9832   Max.   :141388   Max.   : 179.97   Max.   :1252.3   Max.   : 179.986   Max.   :1253.6   Max.   : 179.9374
   R4.PM6.I          R4.PA7.VH          R4.PM7.V          R4.PA8.VH             R4.PM8.V          R4.PA9.VH             R4.PM9.V
Min.   :   0.0   Min.   :-179.98   Min.   : 18270   Min.   :-176.76728   Min.   :   0.000   Min.   :-119.44061   Min.   :   0.000
1st Qu.: 308.0   1st Qu.:-105.59   1st Qu.:131058   1st Qu.:   0.00000   1st Qu.:   0.000   1st Qu.:   0.00000   1st Qu.:   0.000
Median : 380.0   Median : -27.25   Median :131660   Median :   0.00000   Median :   0.000   Median :   0.00000   Median :   0.000
Mean   : 385.3   Mean   : -15.00   Mean   :131372   Mean   :  -0.01219   Mean   :   4.596   Mean   :   0.02692   Mean   :   3.093
3rd Qu.: 452.6   3rd Qu.:  73.79   3rd Qu.:132161   3rd Qu.:   0.00000   3rd Qu.:   0.000   3rd Qu.:   0.00000   3rd Qu.:   0.000
Max.   :1250.5   Max.   : 179.85   Max.   :141112   Max.   : 171.63666   Max.   :7567.229   Max.   : 176.59699   Max.   :7011.945
  R4.PA10.IH         R4.PM10.I         R4.PA11.IH         R4.PM11.I          R4.PA12.IH         R4.PM12.I           R4.F
Min.   :-179.99   Min.   :   0.0   Min.   :-180.00   Min.   :  0.000   Min.   :-179.932   Min.   :  0.000   Min.   :57.08
1st Qu.:-104.52   1st Qu.: 310.6   1st Qu.:   0.00   1st Qu.:  0.000   1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:60.00
Median : -22.44   Median : 381.4   Median :   0.00   Median :  0.000   Median :   0.000   Median :  0.000   Median :60.00
Mean   : -13.43   Mean   : 387.1   Mean   :  -3.78   Mean   :  4.021   Mean   :   5.055   Mean   :  3.880   Mean   :60.00
3rd Qu.:  71.93   3rd Qu.: 454.2   3rd Qu.:   0.00   3rd Qu.:  6.043   3rd Qu.:   0.000   3rd Qu.:  5.875   3rd Qu.:60.00
Max.   : 179.97   Max.   :1252.1   Max.   : 179.98   Max.   :235.067   Max.   : 179.840   Max.   :231.920   Max.   :62.06
     R4.DF              R4.PA.Z             R4.PA.ZH              R4.S        control_panel_log1 control_panel_log2 control_panel_log3
Min.   :-2.570000   Min.   :    2.583   Min.   :-3.093717   Min.   :    0   0:13142            0:13142            0:13142
1st Qu.: 0.000000   1st Qu.:    8.325   1st Qu.:-0.019700   1st Qu.:    0
Median : 0.000000   Median :   10.014   Median : 0.019998   Median :    0
Mean   : 0.001034   Mean   :   26.022   Mean   :-0.004469   Mean   :  432
3rd Qu.: 0.000000   3rd Qu.:   12.537   3rd Qu.: 0.059627   3rd Qu.:    0
Max.   : 1.830000   Max.   :20718.977   Max.   : 3.026952   Max.   :270336
control_panel_log4 relay1_log relay2_log relay3_log relay4_log snort_log1 snort_log2 snort_log3 snort_log4 marker
0:13142            0:12766    0:12757    0:12819    0:12847    0:13140    0:13141    0:13139    0:13140    0:4405
                   1:  376    1:  385    1:  323    1:  295    1:    2    1:    1    1:    3    1:    2    1:8737
```

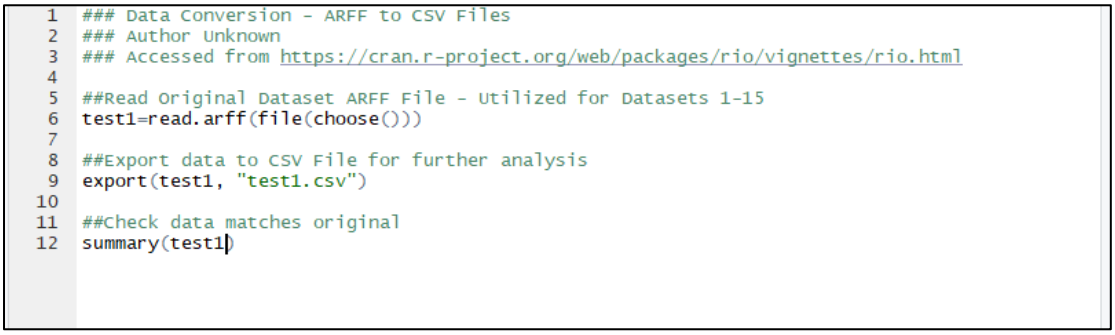
```
> str(aurora, list.len=129)
'data.frame':   13142 obs. of  129 variables:
 $ R1.PA1.VH         : num  70.4 73.7 73.7 74.1 74.6 ...
 $ R1.PM1.V          : num  127673 130281 130306 130582 131083 ...
 $ R1.PA2.VH         : num  -49.6 -46.3 -46.3 -45.9 -45.4 ...
 $ R1.PM2.V          : num  127648 130256 130281 130557 131058 ...
 $ R1.PA3.VH         : num  -170 -166 -166 -166 -165 ...
 $ R1.PM3.V          : num  127723 130356 130381 130657 131158 ...
 $ R1.PA4.IH         : num  65.7 71.8 71.8 72.2 72.1 ...
 $ R1.PM4.I          : num  606 484 484 483 485 ...
 $ R1.PA5.IH         : num  -57 -50.9 -50.9 -50.4 -50 ...
 $ R1.PM5.I          : num  627 501 501 499 498 ...
 $ R1.PA6.IH         : num  -174 -167 -167 -167 -167 ...
 $ R1.PM6.I          : num  602 481 481 481 485 ...
 $ R1.PA7.VH         : num  70.4 73.7 73.8 74.1 74.6 ...
 $ R1.PM7.V          : num  127673 130306 130331 130582 131108 ...
 $ R1.PA8.VH         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ R1.PM8.V          : num  0 0 0 0 0 0 0 0 0 0 ...
 $ R1.PA9.VH         : num  0 0 0 0 0 0 0 0 0 0 ...
 $ R1.PM9.V          : num  0 0 0 0 0 0 0 0 0 0 ...
 $ R1.PA10.IH        : num  65 71.1 71.1 71.5 71.5 ...
 $ R1.PM10.I         : num  612 488 488 488 489 ...
 $ R1.PA11.IH        : num  119 126 125 128 128 ...
 $ R1.PM11.I         : num  13.18 10.62 10.62 9.7 7.51 ...
 $ R1.PA12.IH        : num  -100.9 -95.9 -94.5 -96.7 -99.9 ...
 $ R1.PM12.I         : num  13.92 11.35 11.35 10.44 8.61 ...
 $ R1.F              : num  60 60 60 60 60 ...
 $ R1.DF             : num  0.01 0 0 0 0 0 0 0 0 0 ...
 $ R1.PA.Z           : num  6.39 8.19 8.19 8.17 8.08 ...
 $ R1.PA.ZH          : num  0.0763 0.0249 0.0279 0.0256 0.0329 ...
 $ R1.S              : int  0 0 0 0 0 0 0 0 0 0 ...
 $ R2.PA1.VH         : num  60.7 66.1 66.1 66.5 67 ...
 $ R2.PM1.V          : num  124632 128277 128284 128585 129107 ...
 $ R2.PA2.VH         : num  -59.3 -53.9 -53.9 -53.4 -52.9 ...
 $ R2.PM2.V          : num  124484 128126 128144 128443 128975 ...
 $ R2.PA3.VH         : num  -179 -174 -174 -173 -173 ...
 $ R2.PM3.V          : num  124715 128355 128383 128673 129197 ...
 $ R2.PA4.IH         : num  -120 -115 -115 -115 -115 ...
 $ R2.PM4.I          : num  613 489 489 490 491 ...
 $ R2.PA5.IH         : num  118 122 122 123 123 ...
 $ R2.PM5.I          : num  633 506 506 504 502 ...
 $ R2.PA6.IH         : num  0.86 5.47 5.46 5.69 5.41 ...
 $ R2.PM6.I          : num  610 487 488 487 490 ...
```

```
$ R2.PA7.VH          : num  60.7 66.1 66.1 66.5 67.1 ...
$ R2.PM7.V           : num  124612 128252 128270 128567 129092 ...
$ R2.PA8.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R2.PM8.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R2.PA9.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R2.PM9.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R2.PA10.IH         : num  -120 -116 -116 -116 -115 ...
$ R2.PM10.I          : num  618 494 494 494 494 ...
$ R2.PA11.IH         : num  -64.1 -59.7 -58.7 -61.7 -60.6 ...
$ R2.PM11.I          : num  12.77 10.8 10.4 9.65 7.51 ...
$ R2.PA12.IH         : num  69.4 72.4 72.2 75.5 74 ...
$ R2.PM12.I          : num  12.83 10.81 10.77 9.74 7.53 ...
$ R2.F               : num  60 60 60 60 60 ...
$ R2.DF              : num  0.02 0 0 0 0 0 0 0 0 0 ...
$ R2.PA.Z            : num  6.13 7.92 7.93 7.89 7.79 ...
$ R2.PA.ZH           : num  3.14 -3.14 -3.14 -3.13 -3.11 ...
$ R2.S               : int  0 0 0 0 0 0 0 0 0 0 ...
$ R3.PA1.VH          : num  60.7 66.1 66.1 66.5 67 ...
$ R3.PM1.V           : num  124188 127824 127824 128124 128676 ...
$ R3.PA2.VH          : num  -59.3 -53.9 -53.9 -53.5 -52.9 ...
$ R3.PM2.V           : num  124163 127798 127798 128099 128651 ...
$ R3.PA3.VH          : num  -179 -174 -174 -173 -173 ...
$ R3.PM3.V           : num  124213 127849 127874 128149 128676 ...
$ R3.PA4.IH          : num  -120 -115 -115 -115 -115 ...
$ R3.PM4.I           : num  610 487 487 488 489 ...
$ R3.PA5.IH          : num  118 122 122 123 123 ...
$ R3.PM5.I           : num  628 502 502 501 499 ...
$ R3.PA6.IH          : num  0.659 5.26 5.265 5.472 5.317 ...
$ R3.PM6.I           : num  607 484 484 485 487 ...
$ R3.PA7.VH          : num  60.7 66.1 66.1 66.5 67.1 ...
$ R3.PM7.V           : num  124188 127824 127824 128124 128676 ...
$ R3.PA8.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R3.PM8.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R3.PA9.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R3.PM9.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R3.PA10.IH         : num  -120 -116 -116 -116 -116 ...
$ R3.PM10.I          : num  615 491 491 491 491 ...
$ R3.PA11.IH         : num  -64.8 -60.2 -60.6 -62.4 -61.8 ...
$ R3.PM11.I          : num  12.09 10.25 10.25 9.16 7.51 ...
$ R3.PA12.IH         : num  70.4 73.5 74 76.5 73.4 ...
$ R3.PM12.I          : num  11.9 10.07 9.7 8.97 7.14 ...
$ R3.F               : num  60 60 60 60 60 ...
$ R3.DF              : num  0.02 0 0 0 0 0 0 0 0 0 ...
$ R3.PA.Z            : num  6.11 7.9 7.88 7.86 7.79 ...
```

```
$ R3.PA.ZH           : num  3.14 -3.13 -3.13 -3.12 -3.11 ...
$ R3.S               : int  0 0 0 0 0 0 0 0 0 0 ...
$ R4.PA1.VH          : num  70.5 73.7 73.8 74.1 74.6 ...
$ R4.PM1.V           : num  127723 130331 130356 130632 131133 ...
$ R4.PA2.VH          : num  -49.5 -46.2 -46.2 -45.8 -45.4 ...
$ R4.PM2.V           : num  127096 129704 129729 129980 130506 ...
$ R4.PA3.VH          : num  -170 -166 -166 -166 -165 ...
$ R4.PM3.V           : num  127773 130381 130381 130682 131183 ...
$ R4.PA4.IH          : num  65.6 71.8 71.9 72.2 72.2 ...
$ R4.PM4.I           : num  604 482 481 481 483 ...
$ R4.PA5.IH          : num  -56.9 -50.8 -50.8 -50.3 -49.9 ...
$ R4.PM5.I           : num  622 496 496 495 493 ...
$ R4.PA6.IH          : num  -174 -168 -168 -167 -168 ...
$ R4.PM6.I           : num  600 478 478 478 481 ...
$ R4.PA7.VH          : num  70.5 73.8 73.8 74.2 74.6 ...
$ R4.PM7.V           : num  127523 130130 130155 130431 130933 ...
$ R4.PA8.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R4.PM8.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R4.PA9.VH          : num  0 0 0 0 0 0 0 0 0 0 ...
$ R4.PM9.V           : num  0 0 0 0 0 0 0 0 0 0 ...
$ R4.PA10.IH         : num  65 71.1 71.1 71.5 71.6 ...
$ R4.PM10.I          : num  608 485 485 485 486 ...
$ R4.PA11.IH         : num  119 124 125 126 127 ...
$ R4.PM11.I          : num  12.27 10.25 10.25 9.34 7.32 ...
$ R4.PA12.IH         : num  -102.1 -95.5 -96 -97.3 -101.7 ...
$ R4.PM12.I          : num  11.72 9.7 10.07 9.16 7.14 ...
$ R4.F               : num  60 60 60 60 60 ...
$ R4.DF              : num  0.01 0 0 0 0 0 0 0 0 0 ...
$ R4.PA.Z            : num  6.34 8.14 8.16 8.14 8.04 ...
$ R4.PA.ZH           : num  0.0779 0.0272 0.0267 0.0266 0.0336 ...
$ R4.S               : int  0 0 0 0 0 0 0 0 0 0 ...
$ control_panel_log1: Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
$ control_panel_log2: Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
$ control_panel_log3: Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
$ control_panel_log4: Factor w/ 1 level "0": 1 1 1 1 1 1 1 1 1 1 ...
$ relay1_log         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ relay2_log         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ relay3_log         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ relay4_log         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ snort_log1         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ snort_log2         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ snort_log3         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ snort_log4         : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
$ marker             : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
```

APPENDIX C. EXPORT OF MULTICLASS DATASET

To both modify the original dataset with only applicable data and facilitate initial data analysis in an easily readable format, the following data conversion was conducted via R. This process was iterative, in that each of the fifteen datasets in ARFF format was read in via the file choose function, and then exported to a CSV file and named sequentially (i.e., test1, test2…test15). Upon completion each iteration, the author utilized the summary function to ensure that the data in the newly exported files matched the original datasets, and also checked the CSVs manually via opening the files and observing/comparing the data. A screenshot of the R script is displayed below (Comprehensive R Archive Network, Date Unknown).



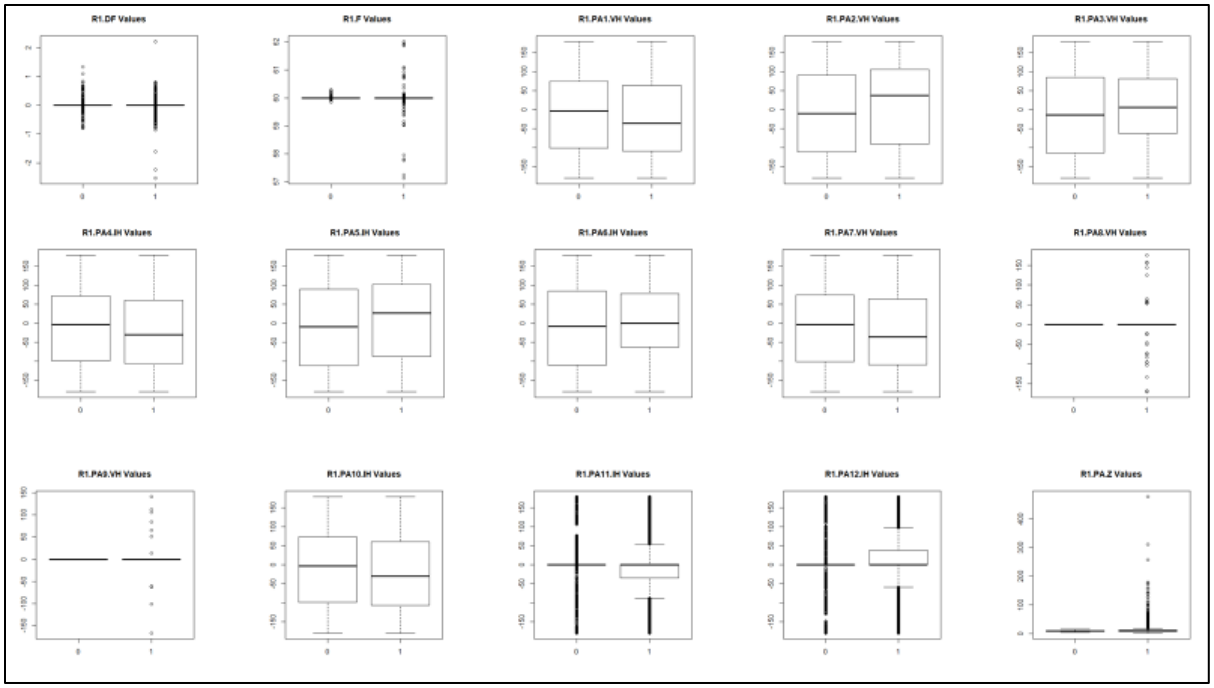
```
### Data Conversion - ARFF to CSV Files
### Author Unknown
### Accessed from https://cran.r-project.org/web/packages/rio/vignettes/rio.html

##Read Original Dataset ARFF File - Utilized for Datasets 1-15
test1=read.arff(file(choose()))

##Export data to CSV File for further analysis
export(test1, "test1.csv")

##Check data matches original
summary(test1)
```

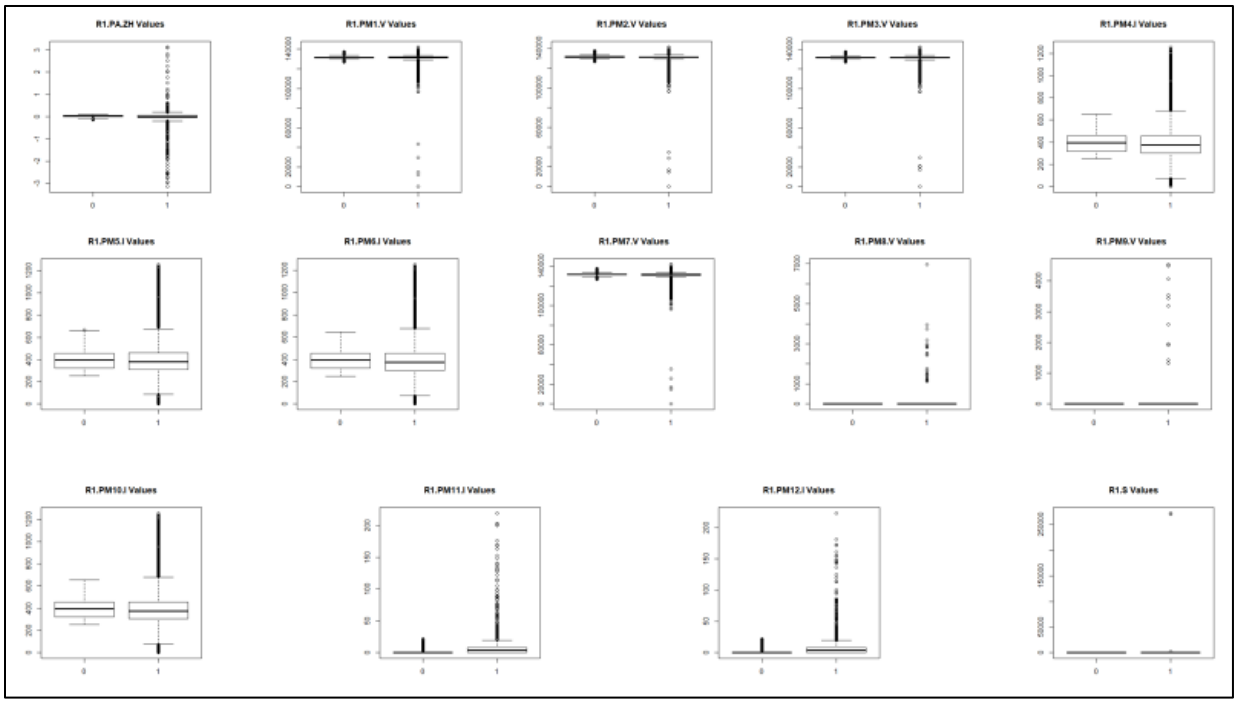
APPENDIX D. INITIAL PLOTS OF AURORA DATASET

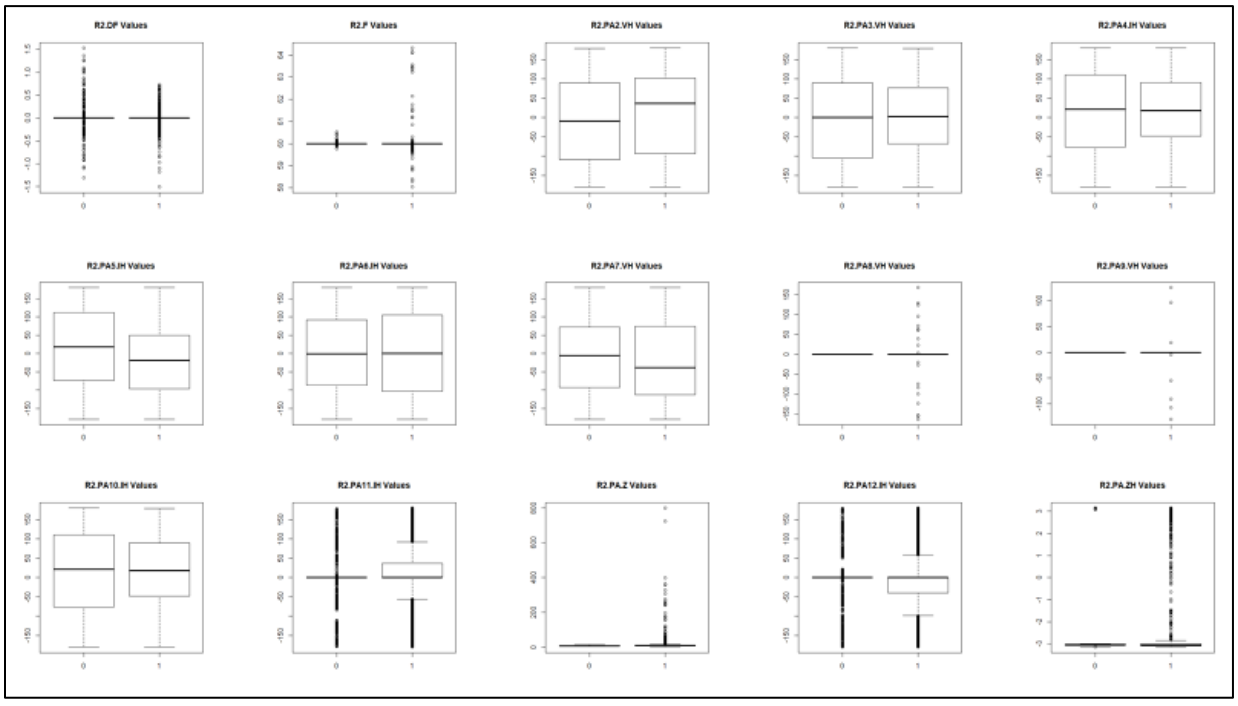
The first R script below was utilized to produce the plots for R1 given the non-standardized dataset. R2, R3, R4 were produced using the same commands (given the different variables with respect to marker). The second R script below was utilized to produce bar charts for R1. Bar charts were selected due to both the marker and relay/SNORT logs were binary data. Relay and snort logs 2-4 were produced utilizing the same commands (control panel graphs omitted due to all data being a 0 indicating normal operations) (Kabacoff, 2017). The plots of each of the 129 variables are listed below the R Script.

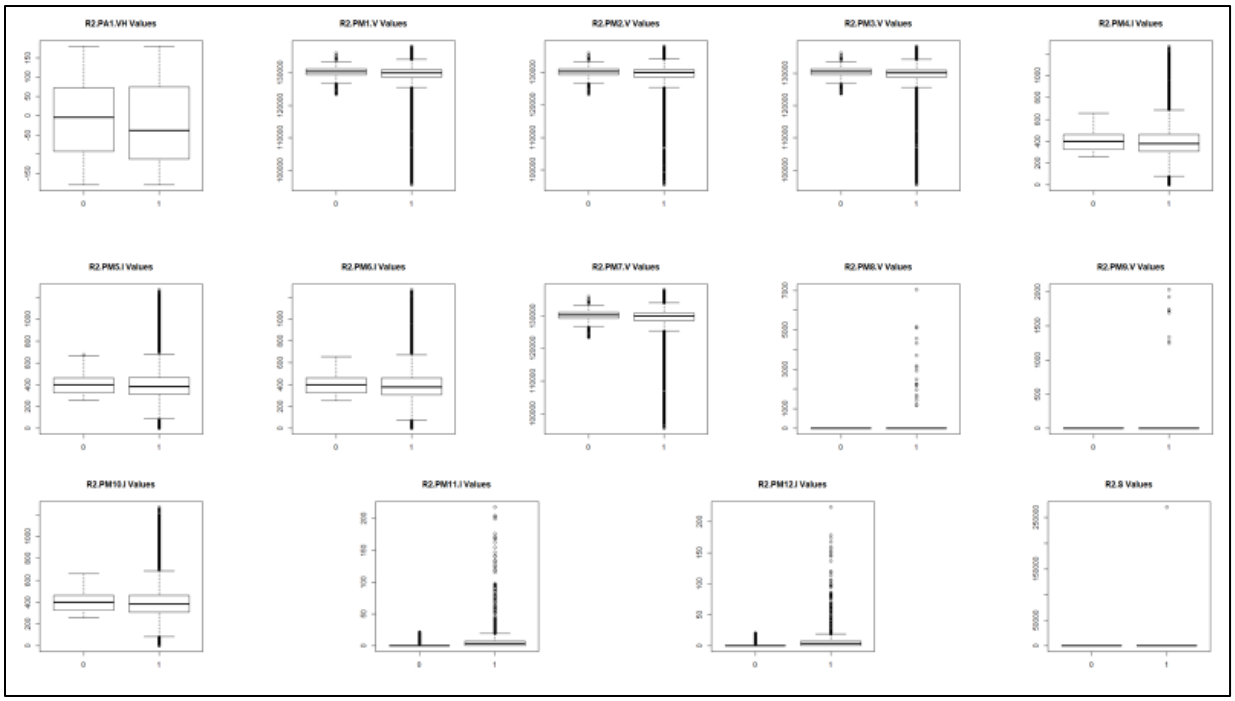
```
## Initial Aurora Data Plots - R1
plot(aurora$marker, aurora$R1.PA1.VH, main="R1.PA1.VH Values")
plot(aurora$marker, aurora$R1.PM1.V, main="R1.PM1.V Values")
plot(aurora$marker, aurora$R1.PA2.VH, main="R1.PA2.VH Values")
plot(aurora$marker, aurora$R1.PM2.V, main="R1.PM2.V Values")
plot(aurora$marker, aurora$R1.PA3.VH, main="R1.PA3.VH Values")
plot(aurora$marker, aurora$R1.PM3.V, main="R1.PM3.V Values")
plot(aurora$marker, aurora$R1.PA4.IH, main="R1.PA4.IH Values")
plot(aurora$marker, aurora$R1.PM4.I, main="R1.PM4.I Values")
plot(aurora$marker, aurora$R1.PA5.IH, main="R1.PA5.IH Values")
plot(aurora$marker, aurora$R1.PM5.I, main="R1.PM5.I Values")
plot(aurora$marker, aurora$R1.PA6.IH, main="R1.PA6.IH Values")
plot(aurora$marker, aurora$R1.PM6.I, main="R1.PM6.I Values")
plot(aurora$marker, aurora$R1.PA7.VH, main="R1.PA7.VH Values")
plot(aurora$marker, aurora$R1.PM7.V, main="R1.PM7.V Values")
plot(aurora$marker, aurora$R1.PA8.VH, main="R1.PA8.VH Values")
plot(aurora$marker, aurora$R1.PM8.V, main="R1.PM8.V Values")
plot(aurora$marker, aurora$R1.PA9.VH, main="R1.PA9.VH Values")
plot(aurora$marker, aurora$R1.PM9.V, main="R1.PM9.V Values")
plot(aurora$marker, aurora$R1.PA10.IH, main="R1.PA10.IH Values")
plot(aurora$marker, aurora$R1.PM10.I, main="R1.PM10.I Values")
plot(aurora$marker, aurora$R1.PA11.IH, main="R1.PA11.IH Values")
plot(aurora$marker, aurora$R1.PM11.I, main="R1.PM11.I Values")
plot(aurora$marker, aurora$R1.PA12.IH, main="R1.PA12.IH Values")
plot(aurora$marker, aurora$R1.PM12.I, main="R1.PM12.I Values")
plot(aurora$marker, aurora$R1.F, main="R1.F Values")
plot(aurora$marker, aurora$R1.DF, main="R1.DF Values")
plot(aurora$marker, aurora$R1.PA.Z, main="R1.PA.Z Values")
plot(aurora$marker, aurora$R1.PA.ZH, main="R1.PA.ZH Values")
plot(aurora$marker, aurora$R1.S, main="R1.S Values")
```

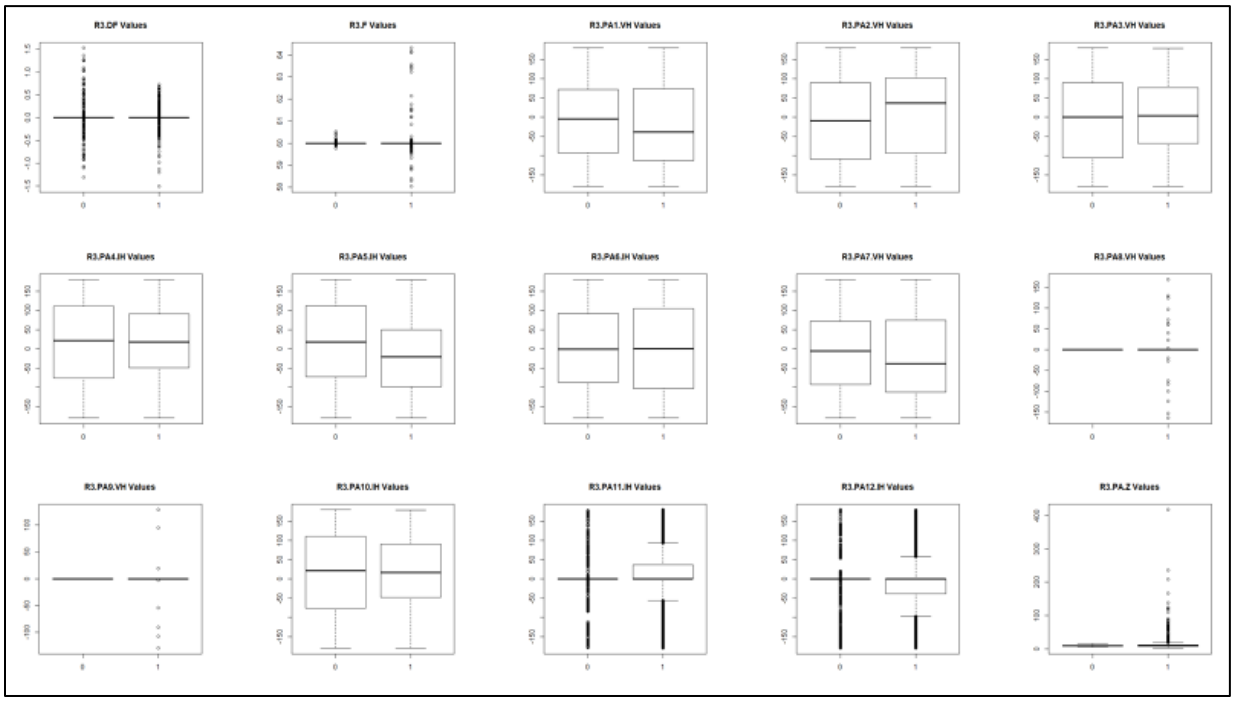
```
340 ##Relay1 Bar Chart
341 ## From https://www.statmethods.net/graphs/bar.html
342 m1r11 <- sum(aurora$marker==1 & aurora$relay1_log==1)
343 m1r10 <- sum(aurora$marker==1 & aurora$relay1_log==0)
344 m0r11 <- sum(aurora$marker==0 & aurora$relay1_log==1)
345 m0r10 <- sum(aurora$marker==0 & aurora$relay1_log==0)
346 relay1plot <- c(m1r11, m1r10, m0r11, m0r10)
347 barplot(relay1plot, names.arg=c("M1,R1", "M1,R0", "M0,R1", "M0,R0"), main="Relay 1 Log Bar Chart")
348
349
```

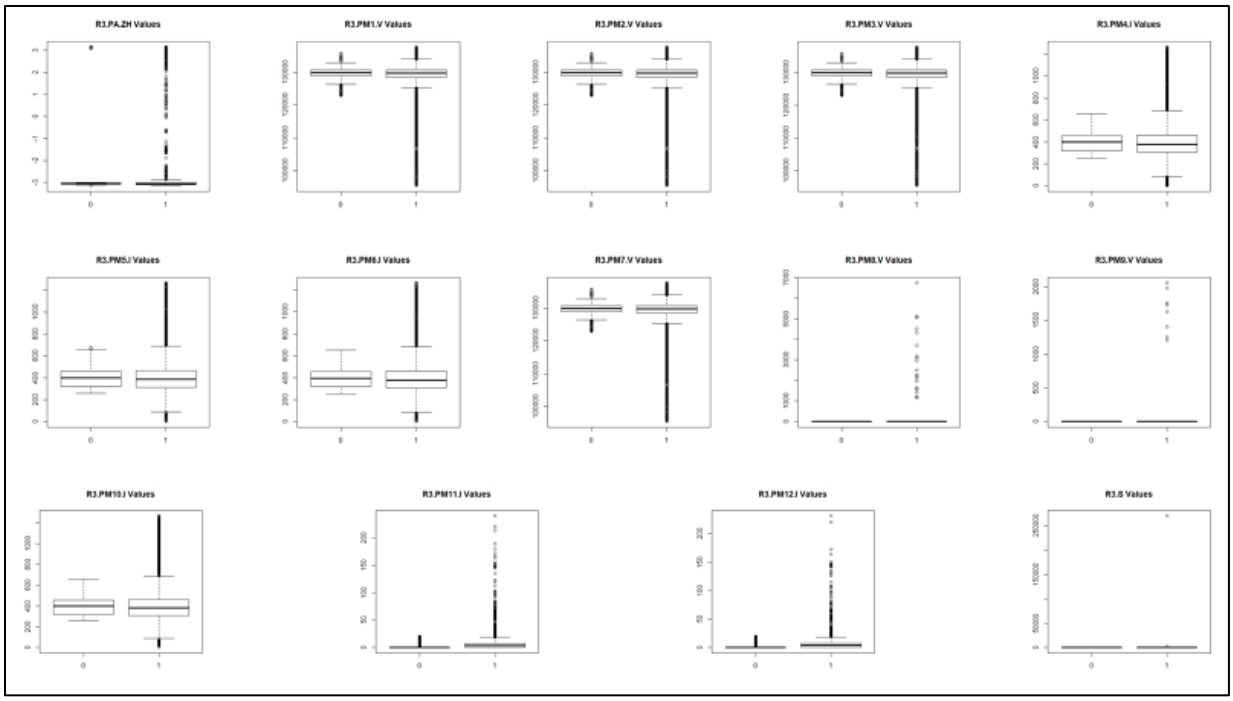


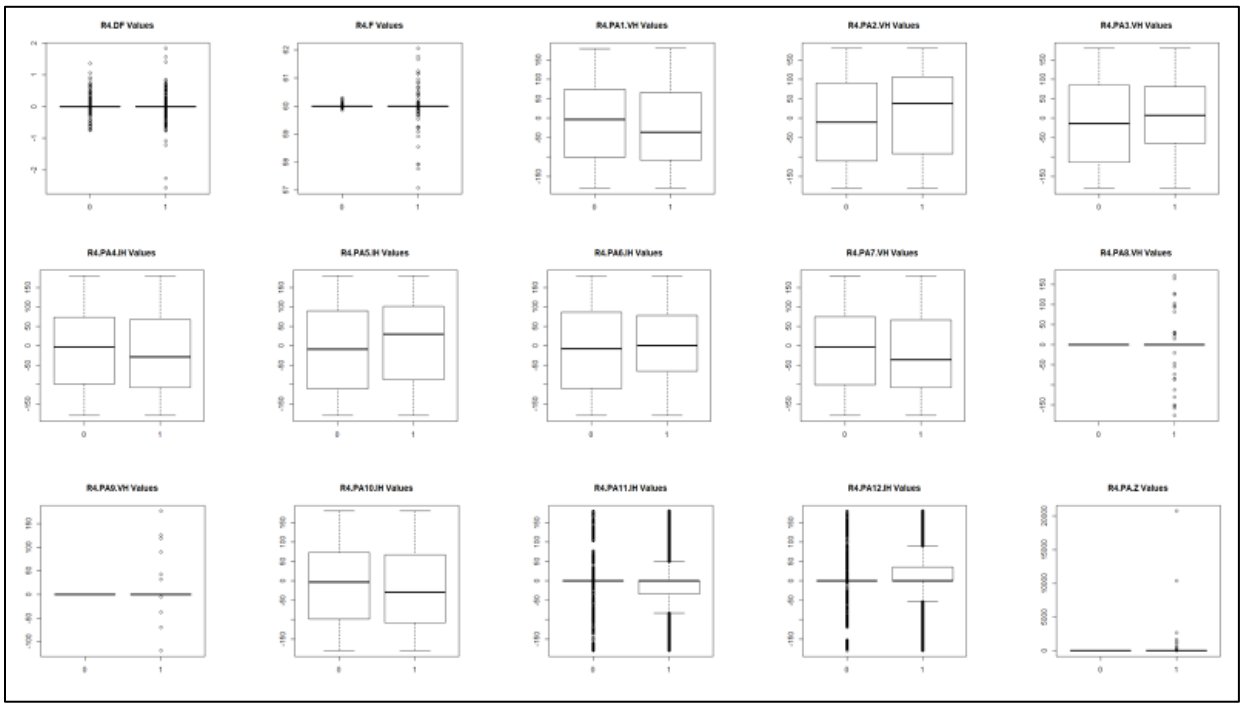

R1.PA.ZH Values
R1.PM1.V Values
R1.PM2.V Values
R1.PM3.V Values
R1.PM4.I Values
R1.PM5.I Values
R1.PM6.I Values
R1.PM7.V Values
R1.PM8.V Values
R1.PM9.V Values
R1.PM10.I Values
R1.PM11.I Values
R1.PM12.I Values
R1.S Values

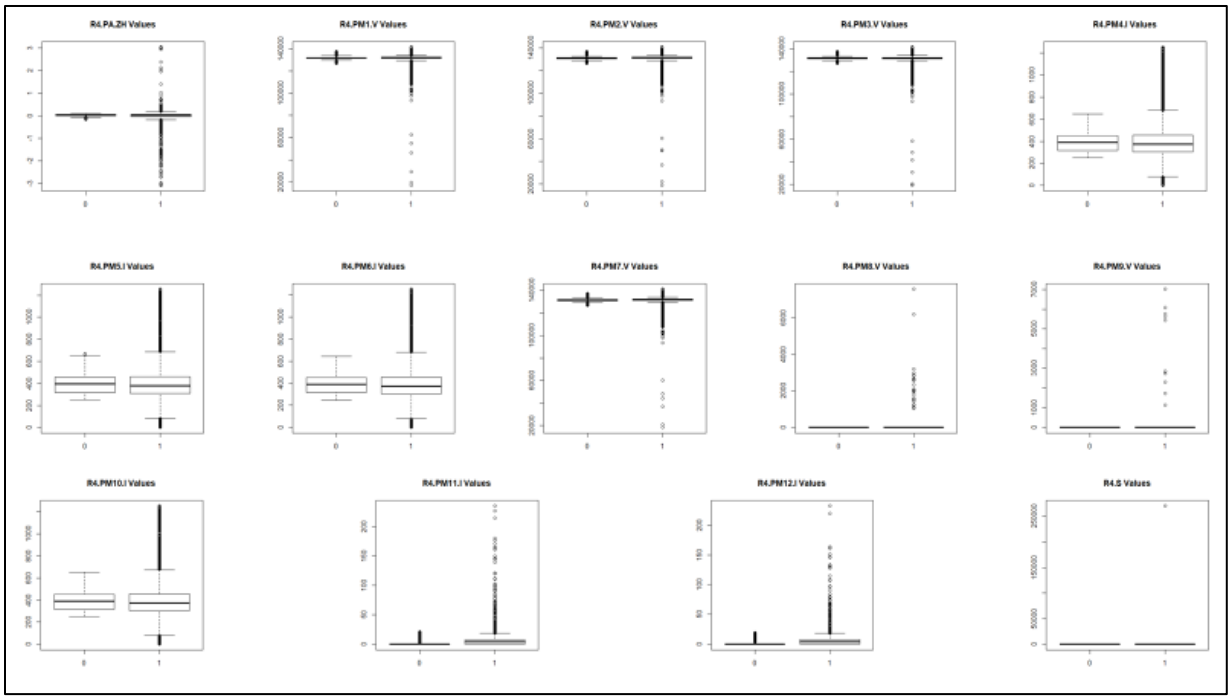

R2.DF Values
R2.F Values
R2.PA2.VH Values
R2.PA3.VH Values
R2.PA4.IH Values
R2.PA5.IH Values
R2.PA6.IH Values
R2.PA7.VH Values
R2.PA8.VH Values
R2.PA9.VH Values
R2.PA10.IH Values
R2.PA11.IH Values
R2.PA.Z Values
R2.PA12.IH Values
R2.PA.ZH Values

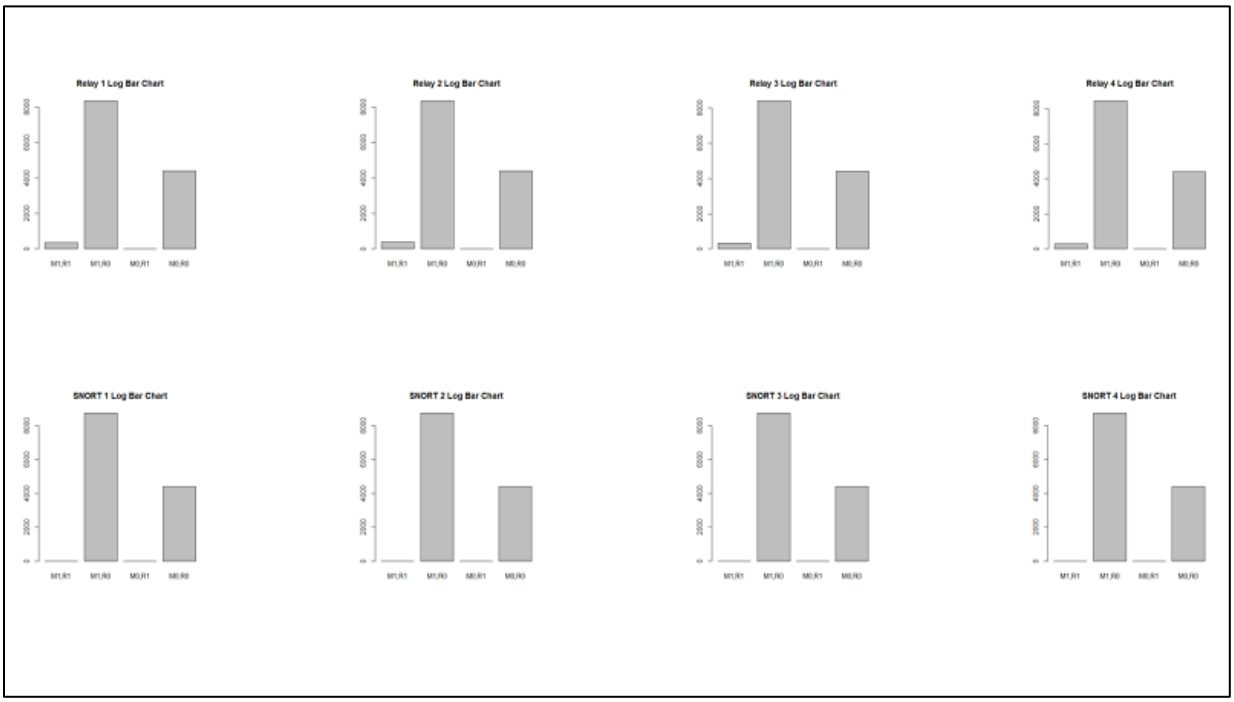

R2.PA1.VH Values
R2.PM1.V Values
R2.PM2.V Values
R2.PM3.V Values
R2.PM4.I Values
R2.PM5.I Values
R2.PM6.I Values
R2.PM7.V Values
R2.PM8.V Values
R2.PM9.V Values
R2.PM10.I Values
R2.PM11.I Values
R2.PM12.I Values
R2.S Values


R3.DF Values
R3.F Values
R3.PA1.VH Values
R3.PA2.VH Values
R3.PA3.VH Values
R3.PA4.IH Values
R3.PA5.IH Values
R3.PA6.IH Values
R3.PA7.VH Values
R3.PA8.VH Values
R3.PA9.VH Values
R3.PA10.IH Values
R3.PA11.IH Values
R3.PA12.IH Values
R3.PA.Z Values


R3.PA.ZH Values
R3.PM1.V Values
R3.PM2.V Values
R3.PM3.V Values
R3.PM4.I Values
R3.PM5.I Values
R3.PM6.I Values
R3.PM7.V Values
R3.PM8.V Values
R3.PM9.V Values
R3.PM10.I Values
R3.PM11.I Values
R3.PM12.I Values
R3.S Values


R4.DF Values
R4.F Values
R4.PA1.VH Values
R4.PA2.VH Values
R4.PA3.VH Values
R4.PA4.IH Values
R4.PA5.IH Values
R4.PA6.IH Values
R4.PA7.VH Values
R4.PA8.VH Values
R4.PA9.VH Values
R4.PA10.IH Values
R4.PA11.IH Values
R4.PA12.IH Values
R4.PA.Z Values


R4.PA.ZH Values
R4.PM1.V Values
R4.PM2.V Values
R4.PM3.V Values
R4.PM4.I Values
R4.PM5.I Values
R4.PM6.I Values
R4.PM7.V Values
R4.PM8.V Values
R4.PM9.V Values
R4.PM10.I Values
R4.PM11.I Values
R4.PM12.I Values
R4.S Values


Relay 1 Log Bar Chart
Relay 2 Log Bar Chart
Relay 3 Log Bar Chart
Relay 4 Log Bar Chart
SNORT 1 Log Bar Chart
SNORT 2 Log Bar Chart
SNORT 3 Log Bar Chart
SNORT 4 Log Bar Chart

APPENDIX E. INITIAL STANDARDIZED LOGISTIC REGRESSION AND STEPWISE LOGISTIC REGRESSION MODELS

The R scripts and subsequent outputs below are of the initial logistic regression model and subsequent stepwise logistic regression model given the standardized dataset (Wiegand, 2018). All logs were removed due to constraints of R to create generalized linear models with numeric data types. The initial output from logistic regression also displays that R4.S should likely be removed due to its importance (all values were 0 for this variable in the non-standardized dataset, and -.04 in the standardized dataset). After the application of the step function, the AIC value was reduced from 254339.8 to 151248.9 with the removal of R4.PM12.I, R1.PM11.I, and R3.PM7.V. For brevity, only the output showing the beginning AIC value and the final results are included below. Residual plots and scripts to identify numbers/characteristics of residuals of both the initial data and subsequent stepwise models are displayed after the R outputs.

```
### - DR WIEGAND OFFICE HOURS 4JUN18 - In relation to Dr. Lathrops question about standardizing data for the SLR
### - Will replicate SWM1 from the SLR at the top of the screen - will compare AIC values and omission of variables

standardizeAll <- as.data.frame(scale(aurora[1:116])) ## Only standardized through the numeric data (column 116)...can't with nominal

## Write/export the StandardizeAll variables and manually add marker so I can include in SLR...
write.csv(standardizeAll, "standardized.csv")

## CHose the standardized dataset...with added nominal values
standardizedAurora <- read.csv(
  file.choose(),
  header = TRUE, sep = ',',
  colClasses=c("marker"="factor", "control_panel_log1"="factor", "control_panel_log2"="factor",
               "control_panel_log3"="factor", "control_panel_log4"="factor", "relay1_log"="factor",
               "relay2_log"="factor", "relay3_log"="factor",
               "relay4_log"="factor", "snort_log1"="factor", "snort_log2"="factor", "snort_log3"="factor", "snort_log4"="factor")
)



### Check data to ensure it's been loaded
summary(standardizedAurora)
```

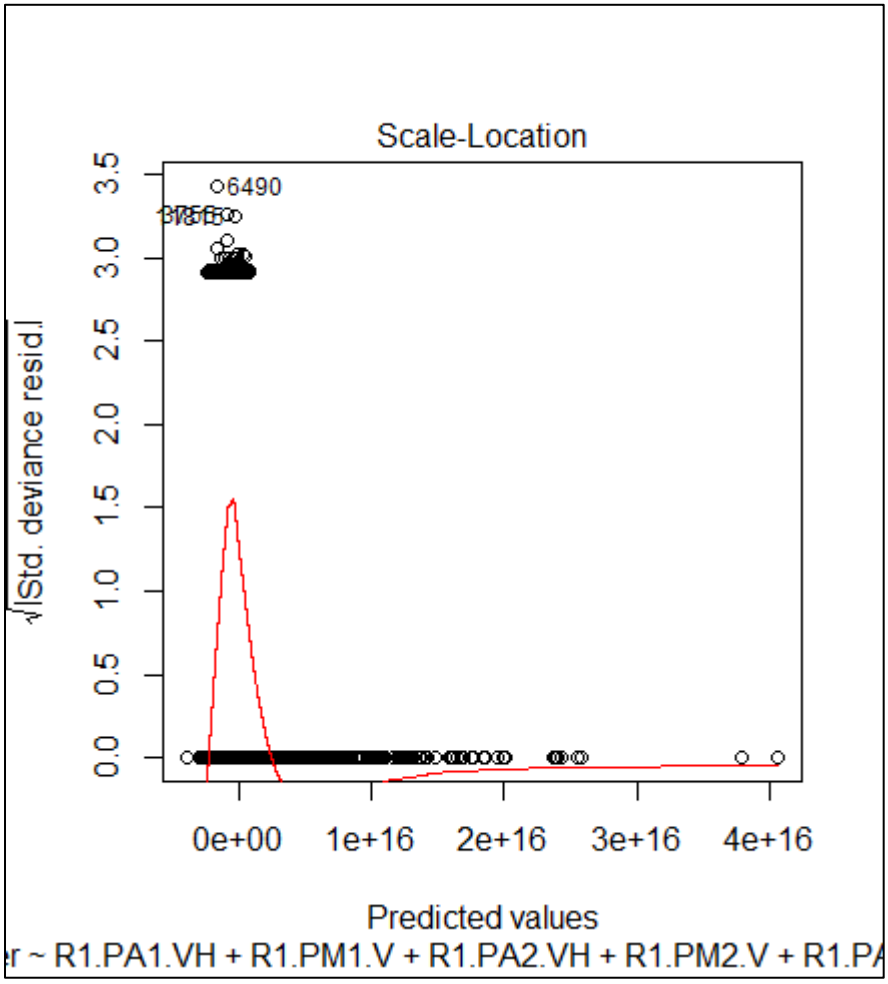
```
standardizedmodel1 <- glm(marker~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH + R1.PM3.V + R1.PA4.IH + R1.PM4.I +
                          R1.PA5.IH + R1.PM5.I + R1.PA6.IH + R1.PM6.I +
                          R1.PA7.VH + R1.PM7.V + R1.PA8.VH + R1.PM8.V +
                          R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
                          R1.PA11.IH + R1.PM11.I + R1.PA12.IH + R1.PM12.I +
                          R1.F + R1.DF + R1.PA.ZH + R1.S +
                          R2.PA1.VH + R2.PM1.V + R2.PA2.VH + R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PA4.IH + R2.PM4.I +
                          R2.PA5.IH + R2.PM5.I + R2.PA6.IH + R2.PM6.I +
                          R2.PA7.VH + R2.PM7.V + R2.PA8.VH + R2.PM8.V +
                          R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I +
                          R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I +
                          R2.F + R2.DF + R2.PA.ZH + R2.S +
                          R3.PA1.VH + R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
                          R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I +
                          R3.PA7.VH + R3.PM7.V + R3.PA8.VH + R3.PM8.V +
                          R3.PA9.VH + R3.PM9.V + R3.PA10.IH + R3.PM10.I +
                          R3.PA11.IH + R3.PM11.I + R3.PA12.IH + R3.PM12.I +
                          R3.F + R3.DF + R3.PA.ZH + R3.S +
                          R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH + R4.PM3.V + R4.PA4.IH + R4.PM4.I +
                          R4.PA5.IH + R4.PM5.I + R4.PA6.IH + R4.PM6.I +
                          R4.PA7.VH + R4.PM7.V + R4.PA8.VH + R4.PM8.V +
                          R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
                          R4.PA11.IH + R4.PM11.I + R4.PA12.IH + R4.PM12.I +
                          R4.F + R4.DF + R4.PA.ZH + R4.S +
                          R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z,
                        #control_panel_log1 + control_panel_log2 + control_panel_log3 +
                        #control_panel_log4 + relay1_log + relay2_log + relay3_log +
                        #relay4_log + snort_log1 + snort_log2 + snort_log3 + snort_log4,
                        data=standardizedAurora, family="binomial")
```

```
Deviance Residuals:
   Min      1Q  Median      3Q     Max
 -8.49    0.00    0.00    8.49    8.49

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error    z value Pr(>|z|)
(Intercept)  7.189e+13  5.854e+05  122811884   <2e-16 ***
R1.PA1.VH    2.328e+15  4.086e+07   56960830   <2e-16 ***
R1.PM1.V    -1.927e+15  1.687e+08  -11424959   <2e-16 ***
R1.PA2.VH   -1.134e+14  2.750e+06  -41239594   <2e-16 ***
R1.PM2.V     9.324e+15  1.737e+08   53684100   <2e-16 ***
R1.PA3.VH    1.489e+14  2.736e+06   54405304   <2e-16 ***
R1.PM3.V     4.695e+15  1.838e+08   25540828   <2e-16 ***
R1.PA4.IH   -3.208e+14  6.311e+06  -50832854   <2e-16 ***
R1.PM4.I    -1.666e+15  1.034e+08  -16109592   <2e-16 ***
R1.PA5.IH    1.343e+14  2.332e+06   57588263   <2e-16 ***
R1.PM5.I    -5.731e+15  9.996e+07  -57338741   <2e-16 ***
R1.PA6.IH   -3.901e+13  2.097e+06  -18602611   <2e-16 ***
R1.PM6.I    -4.284e+15  9.653e+07  -44381110   <2e-16 ***
R1.PA7.VH   -2.371e+15  4.091e+07  -57953611   <2e-16 ***
R1.PM7.V    -1.223e+16  4.951e+08  -24700816   <2e-16 ***
R1.PA8.VH   -3.636e+13  9.370e+05  -38807202   <2e-16 ***
R1.PM8.V    -9.270e+13  1.831e+06  -50630665   <2e-16 ***
R1.PA9.VH    6.085e+13  8.765e+05   69421050   <2e-16 ***
R1.PM9.V    -3.114e+13  1.494e+06  -20840962   <2e-16 ***
R1.PA10.IH   3.644e+14  6.245e+06   58347898   <2e-16 ***
R1.PM10.I    1.119e+16  2.951e+08   37910291   <2e-16 ***
R1.PA11.IH   3.372e+13  1.113e+06   30302724   <2e-16 ***
R1.PM11.I   -8.137e+14  7.121e+06 -114278477   <2e-16 ***
R1.PA12.IH   1.354e+14  1.065e+06  127146998   <2e-16 ***
R1.PM12.I    1.610e+15  6.334e+06  254265391   <2e-16 ***
R1.F        -2.484e+14  2.205e+06 -112679074   <2e-16 ***
R1.DF        8.578e+13  6.763e+05  126829210   <2e-16 ***
R1.PA.ZH    -5.957e+14  9.224e+05 -645787324   <2e-16 ***
R1.S         7.820e+15  3.540e+07  220905747   <2e-16 ***
R2.PA1.VH    3.670e+17  1.199e+10   30621671   <2e-16 ***
R2.PM1.V     1.993e+16  3.000e+08   66427657   <2e-16 ***
R2.PA2.VH   -5.446e+14  2.019e+07  -26975415   <2e-16 ***
R2.PM2.V     3.120e+16  2.918e+08  106907388   <2e-16 ***
R2.PA3.VH    1.311e+14  1.047e+07   12524221   <2e-16 ***
R2.PM3.V     2.640e+16  2.658e+08   99337803   <2e-16 ***
R2.PA4.IH   -2.115e+13  4.799e+06   -4407267   <2e-16 ***
R2.PM4.I    -1.097e+16  1.746e+08  -62835043   <2e-16 ***
R2.PA5.IH   -4.209e+13  1.959e+06  -21481753   <2e-16 ***
```

R2.PM5.I	-1.447e+16	1.698e+08	-85228871	<2e-16	***
R2.PA6.IH	6.484e+13	2.085e+06	31098760	<2e-16	***
R2.PM6.I	-1.239e+16	1.677e+08	-73889744	<2e-16	***
R2.PA7.VH	8.847e+17	1.627e+10	54377403	<2e-16	***
R2.PM7.V	-1.134e+17	6.616e+08	-171441274	<2e-16	***
R2.PA8.VH	5.025e+16	2.805e+08	179164515	<2e-16	***
R2.PM8.V	-3.914e+15	1.027e+08	-38121685	<2e-16	***
R2.PA9.VH	-1.097e+16	1.127e+08	-97309545	<2e-16	***
R2.PM9.V	-3.238e+15	4.581e+07	-70666274	<2e-16	***
R2.PA10.IH	2.107e+14	5.152e+06	40893449	<2e-16	***
R2.PM10.I	3.796e+16	5.083e+08	74689988	<2e-16	***
R2.PA11.IH	4.780e+13	1.186e+06	40315141	<2e-16	***
R2.PM11.I	2.698e+14	8.489e+06	31782948	<2e-16	***
R2.PA12.IH	8.685e+13	1.098e+06	79066788	<2e-16	***
R2.PM12.I	1.182e+14	7.336e+06	16117135	<2e-16	***
R2.F	-1.824e+15	8.286e+07	-22011210	<2e-16	***
R2.DF	-9.630e+13	4.868e+06	-19784151	<2e-16	***
R2.PA.ZH	1.182e+14	9.680e+05	122072421	<2e-16	***
R2.S	-8.571e+15	5.109e+07	-167755530	<2e-16	***
R3.PA1.VH	-3.667e+17	1.199e+10	-30592433	<2e-16	***
R3.PM1.V	2.522e+16	2.257e+08	111736002	<2e-16	***
R3.PA2.VH	6.527e+14	2.012e+07	32441453	<2e-16	***
R3.PM2.V	3.979e+15	2.236e+08	17799499	<2e-16	***
R3.PA3.VH	-5.933e+13	1.053e+07	-5637189	<2e-16	***
R3.PM3.V	1.532e+16	1.964e+08	78001546	<2e-16	***
R3.PA4.IH	-2.479e+14	6.243e+06	-39702590	<2e-16	***
R3.PM4.I	-2.326e+16	1.852e+08	-125597150	<2e-16	***
R3.PA5.IH	-3.699e+13	1.970e+06	-18774466	<2e-16	***
R3.PM5.I	-2.699e+16	1.875e+08	-143985345	<2e-16	***
R3.PA6.IH	1.339e+13	2.082e+06	6431589	<2e-16	***
R3.PM6.I	-2.065e+16	1.889e+08	-109313532	<2e-16	***
R3.PA7.VH	-8.850e+17	1.627e+10	-54391663	<2e-16	***
R3.PM7.V	-9.334e+15	3.597e+08	-25949399	<2e-16	***
R3.PA8.VH	-5.031e+16	2.806e+08	-179270016	<2e-16	***
R3.PM8.V	4.000e+15	1.014e+08	39456351	<2e-16	***
R3.PA9.VH	1.116e+16	1.137e+08	98168637	<2e-16	***
R3.PM9.V	3.153e+15	4.577e+07	68893712	<2e-16	***
R3.PA10.IH	2.778e+13	6.422e+06	4326450	<2e-16	***
R3.PM10.I	7.704e+16	5.567e+08	138391808	<2e-16	***
R3.PA11.IH	-7.481e+13	1.197e+06	-62503649	<2e-16	***
R3.PM11.I	4.084e+14	7.899e+06	51696464	<2e-16	***
R3.PA12.IH	-3.943e+13	1.104e+06	-35712348	<2e-16	***
R3.PM12.I	2.640e+14	7.393e+06	35714205	<2e-16	***
R3.F	1.760e+15	8.285e+07	21246944	<2e-16	***

```
R3.DF         1.155e+14  4.871e+06    23708714    <2e-16 ***
R3.PA.ZH     -3.754e+14  9.683e+05 -387657457    <2e-16 ***
R3.S          8.949e+14  3.677e+07    24336956    <2e-16 ***
R4.PA1.VH    -1.453e+17  1.730e+09   -84002645    <2e-16 ***
R4.PM1.V      2.287e+15  8.928e+07    25616623    <2e-16 ***
R4.PA2.VH     1.605e+13  2.646e+06     6064328    <2e-16 ***
R4.PM2.V      4.577e+14  9.146e+07     5004212    <2e-16 ***
R4.PA3.VH    -9.826e+13  2.782e+06   -35321679    <2e-16 ***
R4.PM3.V      1.480e+15  9.291e+07    15929699    <2e-16 ***
R4.PA4.IH     4.500e+14  6.190e+06    72703844    <2e-16 ***
R4.PM4.I      1.905e+16  1.479e+08   128847684    <2e-16 ***
R4.PA5.IH     4.895e+13  2.345e+06    20875194    <2e-16 ***
R4.PM5.I      1.768e+16  1.517e+08   116491227    <2e-16 ***
R4.PA6.IH     4.765e+13  2.148e+06    22182888    <2e-16 ***
R4.PM6.I      1.763e+16  1.545e+08   114126850    <2e-16 ***
R4.PA7.VH     1.454e+17  1.730e+09    84044784    <2e-16 ***
R4.PM7.V     -3.944e+15  2.678e+08   -14728204    <2e-16 ***
R4.PA8.VH     6.016e+13  1.018e+06    59110215    <2e-16 ***
R4.PM8.V      2.244e+14  1.931e+06   116205109    <2e-16 ***
R4.PA9.VH     1.587e+14  1.053e+06   150701424    <2e-16 ***
R4.PM9.V      7.258e+13  1.256e+06    57785445    <2e-16 ***
R4.PA10.IH   -5.545e+14  6.215e+06   -89215616    <2e-16 ***
R4.PM10.I    -6.020e+16  4.472e+08  -134618756    <2e-16 ***
R4.PA11.IH   -2.786e+12  1.113e+06    -2502950    <2e-16 ***
R4.PM11.I    -8.631e+14  7.507e+06  -114975431    <2e-16 ***
R4.PA12.IH   -1.309e+13  1.070e+06   -12232241    <2e-16 ***
R4.PM12.I     6.851e+14  7.048e+06    97197947    <2e-16 ***
R4.F         -1.858e+14  2.194e+06   -84660656    <2e-16 ***
R4.DF        -3.100e+13  6.836e+05   -45345818    <2e-16 ***
R4.PA.ZH     -4.404e+14  1.161e+06  -379446232    <2e-16 ***
R4.S                 NA         NA          NA         NA
R1.PA.Z      -9.760e+13  1.043e+06   -93584893    <2e-16 ***
R2.PA.Z       2.141e+14  8.850e+05   241869262    <2e-16 ***
R3.PA.Z       9.593e+12  1.106e+06     8674986    <2e-16 ***
R4.PA.Z       2.328e+13  8.433e+05    27609079    <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance:  16764  on 13141  degrees of freedom
Residual deviance: 254108  on 13026  degrees of freedom
AIC: 254340

Number of Fisher Scoring iterations: 25
```

```
plot(standardizedmodel1)

summary(standardizedmodel1)

stdris <- standardizedmodel1$residuals
summary(stdris)


sum(filter(standardizedAurora, stdris>2)$marker==1)
sum(filter(standardizedAurora, stdris>2)$marker==0)

##OUTPUT: 3469 of the residuals over the 2.0 mark (anomalies) from std dataset
##OUPUT: 0 of the residuals over the 2.0 mark from std dataset are normal operations

sum(filter(standardizedAurora, stdris<2)$marker==1)
sum(filter(standardizedAurora, stdris<2)$marker==0)

##OUTPUT: 5268 of the residuals less than the 2.0 mark from std dataset are attacks
##OUPUT: 4405 of the residuals less than the 2.0 mark from std dataset are normal operations

## after ensuring the data is loaded, will use the step function below
```




```
stepwisestdmodel1 <- step(standardizedmodel1, direction="both")
plot(stepwisestdmodel1)
### INITIAL AIC VALUE: 254339.8
### FINAL AIC VALUE: 151248.9 (with removal of R4.PM12.I, R1.PM11.I, and R3.PM7.V)
## Initial AIC Lower, Final AIC lowerr...removal of three variables above, only 1 replicated in initial model w/ non
## standardized data...
```

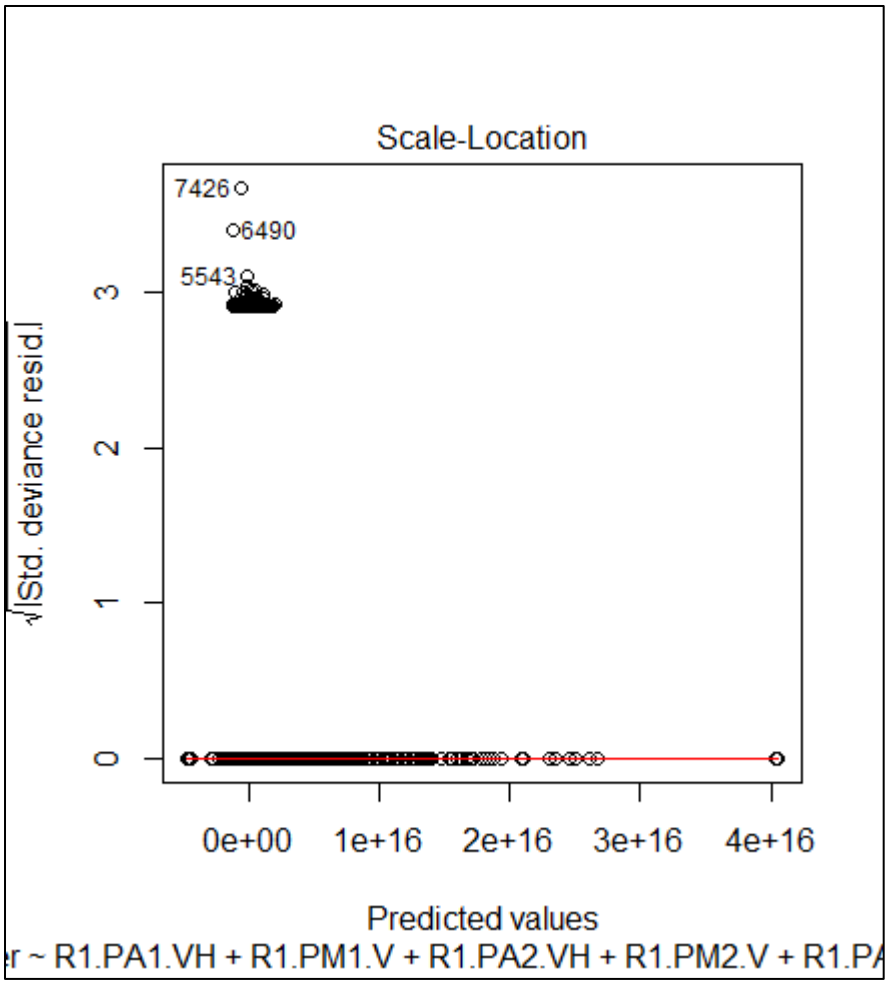
```
Start:  AIC=254339.8
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PM8.V + R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
    R1.PA11.IH + R1.PM11.I + R1.PA12.IH + R1.PM12.I + R1.F +
    R1.DF + R1.PA.ZH + R1.S + R2.PA1.VH + R2.PM1.V + R2.PA2.VH +
    R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PA4.IH + R2.PM4.I +
    R2.PA5.IH + R2.PM5.I + R2.PA6.IH + R2.PM6.I + R2.PA7.VH +
    R2.PM7.V + R2.PA8.VH + R2.PM8.V + R2.PA9.VH + R2.PM9.V +
    R2.PA10.IH + R2.PM10.I + R2.PA11.IH + R2.PM11.I + R2.PA12.IH +
    R2.PM12.I + R2.F + R2.DF + R2.PA.ZH + R2.S + R3.PA1.VH +
    R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH + R3.PM3.V +
    R3.PA4.IH + R3.PM4.I + R3.PA5.IH + R3.PM5.I + R3.PA6.IH +
    R3.PM6.I + R3.PA7.VH + R3.PM7.V + R3.PA8.VH + R3.PM8.V +
    R3.PA9.VH + R3.PM9.V + R3.PA10.IH + R3.PM10.I + R3.PA11.IH +
    R3.PM11.I + R3.PA12.IH + R3.PM12.I + R3.F + R3.DF + R3.PA.ZH +
    R3.S + R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH +
    R4.PM3.V + R4.PA4.IH + R4.PM4.I + R4.PA5.IH + R4.PM5.I +
    R4.PA6.IH + R4.PM6.I + R4.PA7.VH + R4.PM7.V + R4.PA8.VH +
    R4.PM8.V + R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
    R4.PA11.IH + R4.PM11.I + R4.PA12.IH + R4.PM12.I + R4.F +
    R4.DF + R4.PA.ZH + R4.S + R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z
```

```
Step:  AIC=151248.9
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PM8.V + R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
    R1.PA11.IH + R1.PA12.IH + R1.PM12.I + R1.F + R1.DF + R1.PA.ZH +
    R1.S + R2.PA1.VH + R2.PM1.V + R2.PA2.VH + R2.PM2.V + R2.PA3.VH +
    R2.PM3.V + R2.PA4.IH + R2.PM4.I + R2.PA5.IH + R2.PM5.I +
    R2.PA6.IH + R2.PM6.I + R2.PA7.VH + R2.PM7.V + R2.PA8.VH +
    R2.PM8.V + R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I +
    R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I + R2.F +
    R2.DF + R2.PA.ZH + R2.S + R3.PA1.VH + R3.PM1.V + R3.PA2.VH +
    R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
    R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I + R3.PA7.VH +
    R3.PA8.VH + R3.PM8.V + R3.PA9.VH + R3.PM9.V + R3.PA10.IH +
    R3.PM10.I + R3.PA11.IH + R3.PM11.I + R3.PA12.IH + R3.PM12.I +
    R3.F + R3.DF + R3.PA.ZH + R3.S + R4.PA1.VH + R4.PM1.V + R4.PA2.VH +
    R4.PM2.V + R4.PA3.VH + R4.PM3.V + R4.PA4.IH + R4.PM4.I +
    R4.PA5.IH + R4.PM5.I + R4.PA6.IH + R4.PM6.I + R4.PA7.VH +
    R4.PM7.V + R4.PA8.VH + R4.PM8.V + R4.PA9.VH + R4.PM9.V +
    R4.PA10.IH + R4.PM10.I + R4.PA11.IH + R4.PM11.I + R4.PA12.IH +
    R4.F + R4.DF + R4.PA.ZH + R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z

             Df Deviance    AIC
<none>            151023 151249
- R2.PA9.VH   1   151095 151319
- R3.PM3.V    1   154483 154707
- R2.PA10.IH  1   155492 155716
- R2.S        1   155564 155788
- R1.PM7.V    1   155709 155933
- R2.PM3.V    1   159601 159825
- R4.PA9.VH   1   159601 159825
- R1.PA10.IH  1   160178 160402
- R3.PA9.VH   1   161836 162060
- R2.PA.ZH    1   163350 163574
- R3.PA1.VH   1   167243 167467
- R1.PA5.IH   1   167459 167683
- R2.PM7.V    1   169117 169341
- R1.PA2.VH   1   169766 169990
- R1.PA.Z     1   171352 171576
- R3.PM9.V    1   172721 172945
- R3.PM11.I   1   172865 173089
```

```
- R2.F         1   173658 173882
- R1.PA4.IH    1   174451 174675
- R1.PA3.VH    1   179425 179649
- R2.PA5.IH    1   179930 180154
- R4.PM9.V     1   181588 181812
- R1.PA.ZH     1   182741 182965
+ R3.PM7.V     1   182741 182969
- R4.PM8.V     1   183534 183758
- R1.PM1.V     1   184832 185056
- R3.PA12.IH   1   185913 186137
- R3.PM10.I    1   186922 187146
- R3.PA3.VH    1   188076 188300
- R4.DF        1   190815 191039
- R3.PM8.V     1   192545 192769
- R1.PA1.VH    1   195140 195364
- R4.PA12.IH   1   196726 196950
- R1.PM9.V     1   201700 201924
- R4.PM5.I     1   203070 203294
- R3.PM5.I     1   203430 203654
- R1.PM2.V     1   205521 205745
- R2.PA3.VH    1   207972 208196
- R2.PM8.V     1   211360 211584
- R4.PA.Z      1   212369 212593
- R2.PA11.IH   1   212585 212809
- R4.PA1.VH    1   213955 214179
- R4.PA8.VH    1   214316 214540
- R4.PM1.V     1   215685 215909
- R3.PA7.VH    1   216046 216270
- R3.PA10.IH   1   218713 218937
- R1.PM3.V     1   218785 219009
- R2.PA6.IH    1   218785 219009
- R2.PM6.I     1   220443 220667
- R1.PA8.VH    1   221020 221244
- R1.PM4.I     1   222389 222613
- R4.PM6.I     1   222678 222902
- R2.PA4.IH    1   222894 223118
- R3.PA6.IH    1   224624 224848
- R3.PM2.V     1   224840 225064
- R1.PA12.IH   1   225345 225569
- R3.PA2.VH    1   226426 226650
```

- R3.PA11.IH	1	226498	226722
- R3.PA.ZH	1	228012	228236
- R2.PA.Z	1	228373	228597
- R1.PM5.I	1	229166	229390
- R1.PA11.IH	1	229526	229750
- R1.PA6.IH	1	231833	232057
- R3.F	1	232193	232417
- R4.PA4.IH	1	232914	233138
- R3.PM12.I	1	233491	233715
- R3.PA.Z	1	234572	234796
- R3.DF	1	234644	234868
- R2.PA12.IH	1	234716	234940
- R1.PM6.I	1	235149	235373
- R3.PA5.IH	1	235293	235517
- R4.PA6.IH	1	235798	236022
- R4.PM4.I	1	236663	236887
- R3.PM1.V	1	236951	237175
- R4.PM7.V	1	237384	237608
- R4.PM10.I	1	237960	238184
- R4.PA2.VH	1	238104	238328
- R4.PA3.VH	1	238176	238400
- R4.F	1	238537	238761
- R3.PM6.I	1	240195	240419
- R2.PM11.I	1	240411	240635
- R2.PM2.V	1	240555	240779
- R3.S	1	240700	240924
- R1.PM8.V	1	241132	241356
- R2.PA2.VH	1	241276	241500
- R2.PM12.I	1	242141	242365
- R1.PA7.VH	1	242285	242509
- R1.S	1	242502	242726
- R4.PA11.IH	1	242502	242726
- R4.PA10.IH	1	242718	242942
- R2.DF	1	242790	243014

```
- R3.PA4.IH    1   243367 243591
- R3.PM4.I     1   243727 243951
- R1.F         1   244736 244960
- R4.PM2.V     1   245818 246042
- R2.PM10.I    1   246106 246330
- R4.PA5.IH    1   248701 248925
- R2.PA1.VH    1   248773 248997
- R4.PM3.V     1   248917 249141
- R2.PA7.VH    1   249206 249430
- R4.PM11.I    1   252522 252746
- R2.PM4.I     1   252594 252818
+ R1.PM11.I    1   253243 253471
- R1.PM10.I    1   254036 254260
- R2.PM1.V     1   254757 254981
- R1.DF        1   255261 255485
- R3.PA8.VH    1   257208 257432
- R4.PA7.VH    1   258145 258369
- R2.PA8.VH    1   258505 258729
- R1.PA9.VH    1   259298 259522
- R2.PM9.V     1   259875 260099
- R4.PA.ZH     1   260163 260387
- R2.PM5.I     1   260668 260892
- R1.PM12.I    1   262037 262261
+ R4.PM12.I    1   294693 294921
```



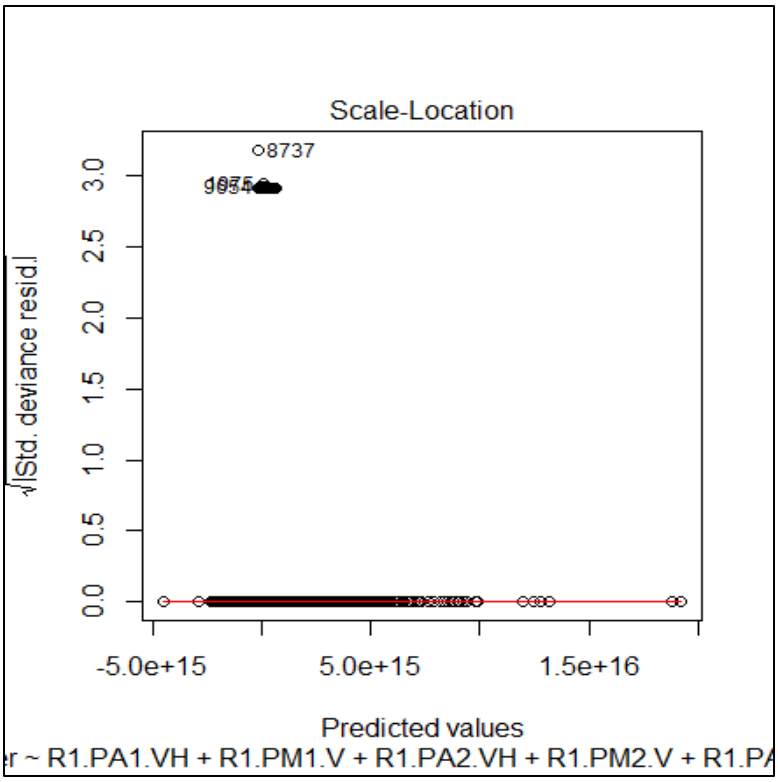

```
1097  stdris2 <- stepwisestdmodel1$residuals
1098  summary(stdris2)
1099
1100
1101  sum(filter(standardizedAurora, stdris2>2)$marker==1)
1102  sum(filter(standardizedAurora, stdris2>2)$marker==0)
1103
1104  ##OUTPUT: 1511 of the residuals over the 2.0 mark (anomalies) from std dataset
1105  ##OUPUT: 0 of the residuals over the 2.0 mark from std dataset are normal operations
1106
1107  sum(filter(standardizedAurora, stdris2<2)$marker==1)
1108  sum(filter(standardizedAurora, stdris2<2)$marker==0)
1109
1110  ##OUTPUT: 7226 of the residuals less than the 2.0 mark from std dataset are attacks
1111  ##OUPUT: 4405 of the residuals less than the 2.0 mark from std dataset are normal operations
```

APPENDIX F. EASY SUBSET STANDADIZED LOGISTIC REGRESSION AND STEPWISE LOGISTIC REGRESSION MODELS

The R scripts/outputs below show the subsetting of the standardized dataset and logistic regression and SLR models given the easy subset (indicated by over a value of 2.0 of the square root of the standard deviance) (Wiegand, 2018). As the results from summary of the easy logistic regression model were replicated from the initial logistic regression in that all variables are significant, it is not included below. After the application of the step function, the AIC value was reduced from 5999.98 to 4333.98 with the removal of R2.PA4.IH.

```
std.easy <- filter(mutate(standardizedAurora, stdris3 = standardizedmodel1$residuals), stdris3 <2.0)
std.hard <- filter(mutate(standardizedAurora, stdris3 = standardizedmodel1$residuals), stdris3 >2.0)
cat(paste("Subset of data with low residual from stepwisemodel1:", dim(std.easy)[1]))
cat(paste("Subset of data with high residual from stepwisemodel1:", dim(std.hard)[1]))
##OUTPUT: 9673 - low residual
##OUTPUT: 3469 - high residual

easystdstepwisemodel1 <- glm(marker~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH + R1.PM3.V + R1.PA4.IH + R1.PM4.I +
                               R1.PA5.IH + R1.PM5.I + R1.PA6.IH + R1.PM6.I +
                               R1.PA7.VH + R1.PM7.V + R1.PA8.VH + R1.PM8.V +
                               R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
                               R1.PA11.IH + R1.PM11.I + R1.PA12.IH + R1.PM12.I +
                               R1.F + R1.DF + R1.PA.ZH + R1.S +
                               R2.PA1.VH + R2.PM1.V + R2.PA2.VH + R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PA4.IH + R2.PM4.I +
                               R2.PA5.IH + R2.PM5.I + R2.PA6.IH + R2.PM6.I +
                               R2.PA7.VH + R2.PM7.V + R2.PA8.VH + R2.PM8.V +
                               R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I +
                               R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I +
                               R2.F + R2.DF + R2.PA.ZH + R2.S +
                               R3.PA1.VH + R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
                               R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I +
                               R3.PA7.VH + R3.PM7.V + R3.PA8.VH + R3.PM8.V +
                               R3.PA9.VH + R3.PM9.V + R3.PA10.IH + R3.PM10.I +
                               R3.PA11.IH + R3.PM11.I + R3.PA12.IH + R3.PM12.I +
                               R3.F + R3.DF + R3.PA.ZH + R3.S +
                               R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH + R4.PM3.V + R4.PA4.IH + R4.PM4.I +
                               R4.PA5.IH + R4.PM5.I + R4.PA6.IH + R4.PM6.I +
                               R4.PA7.VH + R4.PM7.V + R4.PA8.VH + R4.PM8.V +
                               R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
                               R4.PA11.IH + R4.PM11.I + R4.PA12.IH + R4.PM12.I +
                               R4.F + R4.DF + R4.PA.ZH + R4.S +
                               R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z,
                             #control_panel_log1 + control_panel_log2 + control_panel_log3 +
                             #control_panel_log4 + relay1_log + relay2_log + relay3_log +
                             #relay4_log + snort_log1 + snort_log2 + snort_log3 + snort_log4,
                             data=std.easy, family="binomial")
```

```
plot(easystdstepwisemodel1)

easystdris <- easystdstepwisemodel1$residuals
summary(easystdris)


sum(filter(std.easy, easystdris>2)$marker==1)
sum(filter(std.easy, easystdris>2)$marker==0)

##OUTPUT: 9 of the residuals over the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 0 of the residuals over the 2.0 mark from std easy dataset are normal operations

sum(filter(std.easy, easystdris<2)$marker==1)
sum(filter(std.easy, easystdris<2)$marker==0)

##OUTPUT: 5259 of the residuals over the 2.0 mark (anomalies) from stdeasy dataset are attacks
##OUPUT: 4405 of the residuals over the 2.0 mark from std easy dataset are normal operations
```




```
1183  easystdstepwisemodel2 <- step(easystdstepwisemodel1, direction="both")
1184  ### INITIAL AIC VALUE: 5998.98
1185  ### FINAL AIC VALUE: 4338.98 (with removal of R2.PA4.IH )
1186
1187  plot(easystdstepwisemodel2)
```

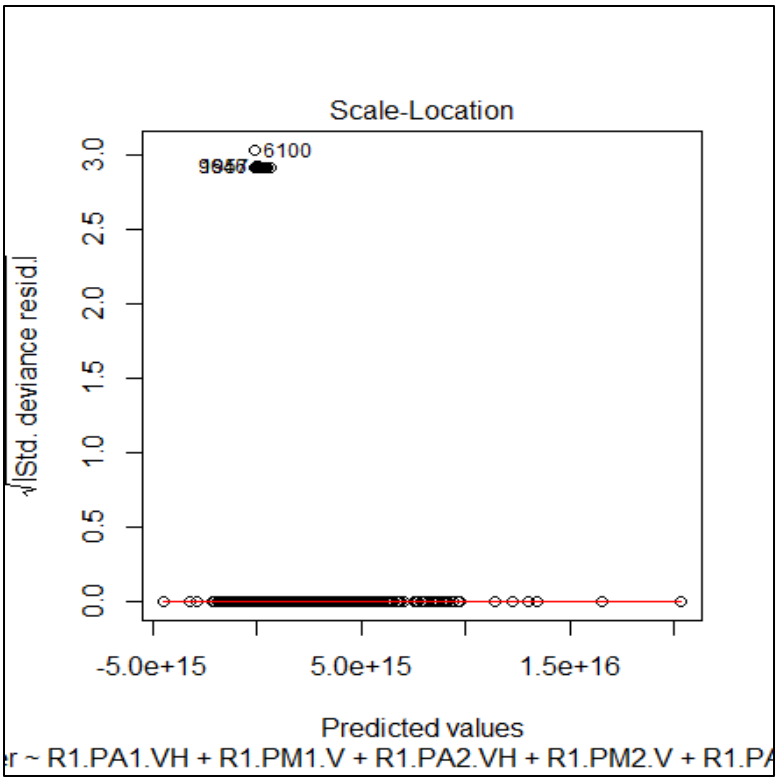
```
Start:  AIC=5998.98
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PM8.V + R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
    R1.PA11.IH + R1.PM11.I + R1.PA12.IH + R1.PM12.I + R1.F +
    R1.DF + R1.PA.ZH + R1.S + R2.PA1.VH + R2.PM1.V + R2.PA2.VH +
    R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PA4.IH + R2.PM4.I +
    R2.PA5.IH + R2.PM5.I + R2.PA6.IH + R2.PM6.I + R2.PA7.VH +
    R2.PM7.V + R2.PA8.VH + R2.PM8.V + R2.PA9.VH + R2.PM9.V +
    R2.PA10.IH + R2.PM10.I + R2.PA11.IH + R2.PM11.I + R2.PA12.IH +
    R2.PM12.I + R2.F + R2.DF + R2.PA.ZH + R2.S + R3.PA1.VH +
    R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH + R3.PM3.V +
    R3.PA4.IH + R3.PM4.I + R3.PA5.IH + R3.PM5.I + R3.PA6.IH +
    R3.PM6.I + R3.PA7.VH + R3.PM7.V + R3.PA8.VH + R3.PM8.V +
    R3.PA9.VH + R3.PM9.V + R3.PA10.IH + R3.PM10.I + R3.PA11.IH +
    R3.PM11.I + R3.PA12.IH + R3.PM12.I + R3.F + R3.DF + R3.PA.ZH +
    R3.S + R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH +
    R4.PM3.V + R4.PA4.IH + R4.PM4.I + R4.PA5.IH + R4.PM5.I +
    R4.PA6.IH + R4.PM6.I + R4.PA7.VH + R4.PM7.V + R4.PA8.VH +
    R4.PM8.V + R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
    R4.PA11.IH + R4.PM11.I + R4.PA12.IH + R4.PM12.I + R4.F +
    R4.DF + R4.PA.ZH + R4.S + R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z
```

```
Step:  AIC=4338.98
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PM8.V + R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
    R1.PA11.IH + R1.PM11.I + R1.PA12.IH + R1.PM12.I + R1.F +
    R1.DF + R1.PA.ZH + R1.S + R2.PA1.VH + R2.PM1.V + R2.PA2.VH +
    R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PM4.I + R2.PA5.IH +
    R2.PM5.I + R2.PA6.IH + R2.PM6.I + R2.PA7.VH + R2.PM7.V +
    R2.PA8.VH + R2.PM8.V + R2.PA9.VH + R2.PM9.V + R2.PA10.IH +
    R2.PM10.I + R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I +
    R2.F + R2.DF + R2.PA.ZH + R2.S + R3.PA1.VH + R3.PM1.V + R3.PA2.VH +
    R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
    R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I + R3.PA7.VH +
    R3.PM7.V + R3.PA8.VH + R3.PM8.V + R3.PA9.VH + R3.PM9.V +
    R3.PA10.IH + R3.PM10.I + R3.PA11.IH + R3.PM11.I + R3.PA12.IH +
    R3.PM12.I + R3.F + R3.DF + R3.PA.ZH + R3.S + R4.PA1.VH +
    R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH + R4.PM3.V +
    R4.PA4.IH + R4.PM4.I + R4.PA5.IH + R4.PM5.I + R4.PA6.IH +
    R4.PM6.I + R4.PA7.VH + R4.PM7.V + R4.PA8.VH + R4.PM8.V +
    R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I + R4.PA11.IH +
    R4.PM11.I + R4.PA12.IH + R4.PM12.I + R4.F + R4.DF + R4.PA.ZH +
    R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z

             Df Deviance     AIC
<none>            4109.0  4339.0
- R2.PM5.I    1   4253.2  4481.2
- R2.PA2.VH   1   4397.3  4625.3
- R2.PM4.I    1   4397.3  4625.3
- R4.PM12.I   1   4469.4  4697.4
- R2.PA3.VH   1   4613.6  4841.6
- R4.PA9.VH   1   4685.7  4913.7
- R1.PM1.V    1   4757.8  4985.8
- R3.PA2.VH   1   4757.8  4985.8
- R3.S        1   4757.8  4985.8
- R1.PA1.VH   1   4829.8  5057.8
- R4.PA3.VH   1   4829.8  5057.8
- R2.PM11.I   1   4974.0  5202.0
- R2.PA1.VH   1   5118.2  5346.2
```

-	R2.PM12.I	1	5118.2	5346.2
-	R2.PA.Z	1	5118.2	5346.2
-	R4.PA7.VH	1	5190.3	5418.3
-	R2.F	1	5262.4	5490.4
-	R1.PM6.I	1	5334.5	5562.5
-	R1.PA9.VH	1	5334.5	5562.5
-	R4.PA1.VH	1	5334.5	5562.5
-	R2.PM3.V	1	5478.6	5706.6
-	R1.PM3.V	1	5550.7	5778.7
-	R2.PM7.V	1	5550.7	5778.7
-	R2.PM10.I	1	5550.7	5778.7
-	R2.S	1	5550.7	5778.7
-	R4.PM11.I	1	5550.7	5778.7
-	R1.PM4.I	1	5622.8	5850.8
-	R1.PA6.IH	1	5622.8	5850.8
-	R3.PA6.IH	1	5622.8	5850.8
-	R3.PA8.VH	1	5622.8	5850.8
-	R4.PM2.V	1	5622.8	5850.8
-	R1.PM10.I	1	5694.9	5922.9
-	R2.DF	1	5767.0	5995.0
+	R2.PA4.IH	1	5767.0	5999.0
-	R2.PA7.VH	1	5839.1	6067.1
-	R4.PM8.V	1	5839.1	6067.1
-	R2.PA9.VH	1	5911.2	6139.2
-	R3.PA5.IH	1	5983.2	6211.2
-	R1.PA5.IH	1	6055.3	6283.3
-	R1.PA.ZH	1	6055.3	6283.3
-	R4.PM6.I	1	6055.3	6283.3
-	R4.PA.ZH	1	6127.4	6355.4
-	R2.PA11.IH	1	6199.5	6427.5
-	R3.PM4.I	1	6199.5	6427.5
-	R3.PM7.V	1	6199.5	6427.5
-	R3.PM10.I	1	6199.5	6427.5
-	R4.F	1	6199.5	6427.5
-	R2.PA6.IH	1	6271.6	6499.6
-	R3.PA3.VH	1	6271.6	6499.6
-	R1.PA4.IH	1	6343.7	6571.7
-	R3.PA7.VH	1	6343.7	6571.7

- R2.PA12.IH	1	6415.8	6643.8
- R4.DF	1	6415.8	6643.8
- R1.PA7.VH	1	6487.9	6715.9
- R1.PA.Z	1	6487.9	6715.9
- R2.PA10.IH	1	6704.1	6932.1
- R1.PA11.IH	1	6776.2	7004.2
- R3.PA12.IH	1	6776.2	7004.2
- R3.PM12.I	1	6776.2	7004.2
- R1.PM2.V	1	6848.3	7076.3
- R1.PM5.I	1	6848.3	7076.3
- R1.PM8.V	1	6848.3	7076.3
- R3.PA4.IH	1	6848.3	7076.3
- R4.PA11.IH	1	6848.3	7076.3
- R4.PM4.I	1	6920.4	7148.4
- R3.PA10.IH	1	6992.5	7220.5
- R3.DF	1	6992.5	7220.5
- R2.PA5.IH	1	7064.6	7292.6
- R3.PA9.VH	1	7064.6	7292.6
- R1.S	1	7136.6	7364.6
- R2.PA8.VH	1	7136.6	7364.6
- R3.PM2.V	1	7136.6	7364.6
- R3.PA.Z	1	7136.6	7364.6
- R1.PA3.VH	1	7208.7	7436.7
- R3.PA11.IH	1	7208.7	7436.7
- R3.PM11.I	1	7425.0	7653.0
- R4.PM3.V	1	7425.0	7653.0
- R4.PA8.VH	1	7497.1	7725.1
- R4.PM7.V	1	7569.2	7797.2
- R3.PM1.V	1	7641.3	7869.3

- R4.PA5.IH	1	7713.3	7941.3
- R4.PM5.I	1	7713.3	7941.3
- R4.PM9.V	1	7713.3	7941.3
- R1.PA10.IH	1	7857.5	8085.5
- R2.PM6.I	1	7857.5	8085.5
- R3.PM6.I	1	7857.5	8085.5
- R1.PM7.V	1	7929.6	8157.6
- R3.F	1	8001.7	8229.7
- R3.PA1.VH	1	8073.8	8301.8
- R1.PA12.IH	1	8145.9	8373.9
- R4.PM10.I	1	8145.9	8373.9
- R1.DF	1	8218.0	8446.0
- R2.PM2.V	1	8290.0	8518.0
- R4.PA10.IH	1	8290.0	8518.0
- R3.PM3.V	1	8434.2	8662.2
- R1.PM12.I	1	8506.3	8734.3
- R4.PA4.IH	1	8506.3	8734.3
- R2.PM1.V	1	8578.4	8806.4
- R1.F	1	8650.5	8878.5
- R2.PA.ZH	1	8866.7	9094.7
- R4.PM1.V	1	9227.2	9455.2
- R3.PM5.I	1	9443.4	9671.4
- R1.PM9.V	1	9515.5	9743.5
- R4.PA12.IH	1	9587.6	9815.6
- R1.PA2.VH	1	9659.7	9887.7
- R3.PA.ZH	1	9731.8	9959.8
- R2.PM9.V	1	10092.2	10320.2
- R4.PA.Z	1	10164.3	10392.3
- R2.PM8.V	1	10308.5	10536.5
- R3.PM8.V	1	10380.6	10608.6
- R4.PA6.IH	1	10524.7	10752.7
- R1.PA8.VH	1	11101.4	11329.4
- R3.PM9.V	1	11461.9	11689.9
- R4.PA2.VH	1	11894.4	12122.4
- R1.PM11.I	1	12182.8	12410.8

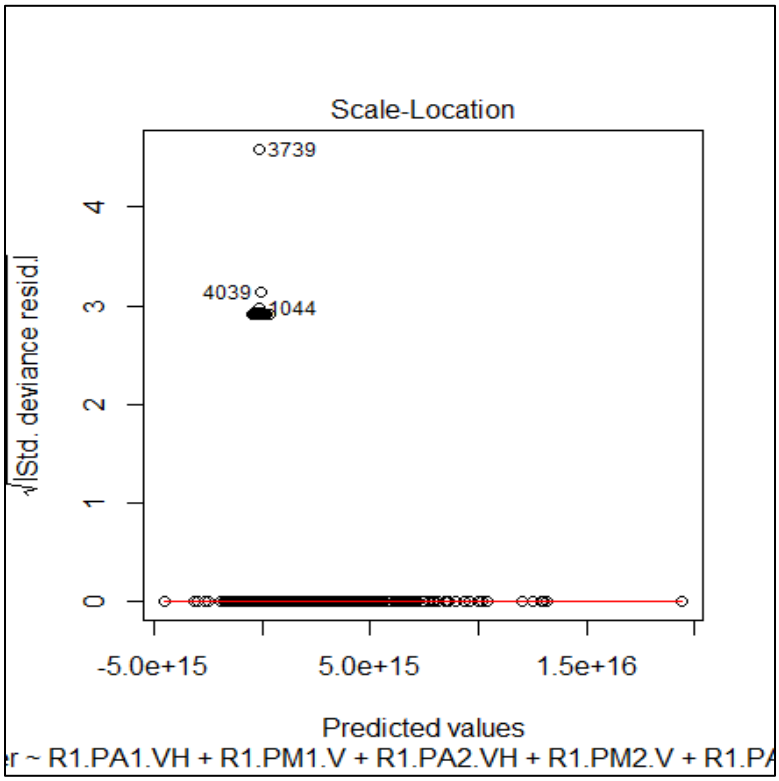



```
easystdris2 <- easystdstepwisemodel2$residuals
summary(easystdris2)


sum(filter(std.easy, easystdris2>2)$marker==1)
sum(filter(std.easy, easystdris2>2)$marker==0)

##OUTPUT: 10 of the residuals over the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 0 of the residuals over the 2.0 mark from std easy dataset are normal operations

sum(filter(std.easy, easystdris2<2)$marker==1)
sum(filter(std.easy, easystdris2<2)$marker==0)

##OUTPUT: 5258 of the residuals less than the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 4405 of the residuals less than the 2.0 mark from std easy dataset are normal operations
```

APPENDIX G. EASY STANDARDIZED STEPWISE MODEL 3 AND 4 LOGISTIC REGRESSION

The same process was utilized for these logistic and stepwise logistic regression models given the easy subset, but with the omission of variables from the initial stepwise logistic regression model (R4.PM12.I, R1.PM11.I, and R3.PM7.V). After the application of the step function, the AIC value was reduced from 10462.4 to 4615.3 with the removal of R2.PA2.VH, R1.PM8.V, R4.PA12.IH, and R3.PA9.VH.

```
### REMOVED R4.PM12.I, R1.PM11.I, and R3.PM7.V
easystdstepwisemodel3 <- glm(marker~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH + R1.PM3.V + R1.PA4.IH + R1.PM4.I +
                               R1.PA5.IH + R1.PM5.I + R1.PA6.IH + R1.PM6.I +
                               R1.PA7.VH + R1.PM7.V + R1.PA8.VH + R1.PM8.V +
                               R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
                               R1.PA11.IH + R1.PA12.IH + R1.PM12.I +
                               R1.F + R1.DF + R1.PA.ZH + R1.S +
                               R2.PA1.VH + R2.PM1.V + R2.PA2.VH + R2.PM2.V + R2.PA3.VH + R2.PM3.V + R2.PA4.IH + R2.PM4.I +
                               R2.PA5.IH + R2.PM5.I + R2.PA6.IH + R2.PM6.I +
                               R2.PA7.VH + R2.PM7.V + R2.PA8.VH + R2.PM8.V +
                               R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I +
                               R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I +
                               R2.F + R2.DF + R2.PA.ZH + R2.S +
                               R3.PA1.VH + R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
                               R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I +
                               R3.PA7.VH + R3.PA8.VH + R3.PM8.V +
                               R3.PA9.VH + R3.PM9.V + R3.PA10.IH + R3.PM10.I +
                               R3.PA11.IH + R3.PM11.I + R3.PA12.IH + R3.PM12.I +
                               R3.F + R3.DF + R3.PA.ZH + R3.S +
                               R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH + R4.PM3.V + R4.PA4.IH + R4.PM4.I +
                               R4.PA5.IH + R4.PM5.I + R4.PA6.IH + R4.PM6.I +
                               R4.PA7.VH + R4.PM7.V + R4.PA8.VH + R4.PM8.V +
                               R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
                               R4.PA11.IH + R4.PM11.I + R4.PA12.IH +
                               R4.F + R4.DF + R4.PA.ZH + R4.S +
                               R1.PA.Z + R2.PA.Z + R3.PA.Z + R4.PA.Z,
                             #control_panel_log1 + control_panel_log2 + control_panel_log3 +
                             #control_panel_log4 + relay1_log + relay2_log + relay3_log +
                             #relay4_log + snort_log1 + snort_log2 + snort_log3 + snort_log4,
                             data=std.easy, family="binomial")

## warnings for the code above:
## Warning messages: glm.fit: fitted probabilities numerically 0 or 1 occurred

summary(easystdstepwisemodel3)

plot(easystdstepwisemodel3)
```



```
easystdris3 <- easystdstepwisemodel3$residuals
summary(easystdris3)


sum(filter(std.easy, easystdris3>2)$marker==1)
sum(filter(std.easy, easystdris3>2)$marker==0)

##OUTPUT: 132 of the residuals over the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 0 of the residuals over the 2.0 mark from std easy dataset are normal operations

sum(filter(std.easy, easystdris3<2)$marker==1)
sum(filter(std.easy, easystdris3<2)$marker==0)

##OUTPUT: 5136 of the residuals less than the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 4405 of the residuals less than the 2.0 mark from std easy dataset are normal operations
```

```
easystdstepwisemodel4 <- step(easystdstepwisemodel3, direction="both")

### INITIAL AIC VALUE: 10462.4
### FINAL AIC VALUE: 4615.3 (with removal of R2.PA2.VH, R1.PM8.V, R4.PA12.IH, and R3.PA9.VH)

summary(easystdstepwisemodel4)

plot(easystdstepwisemodel4)
```

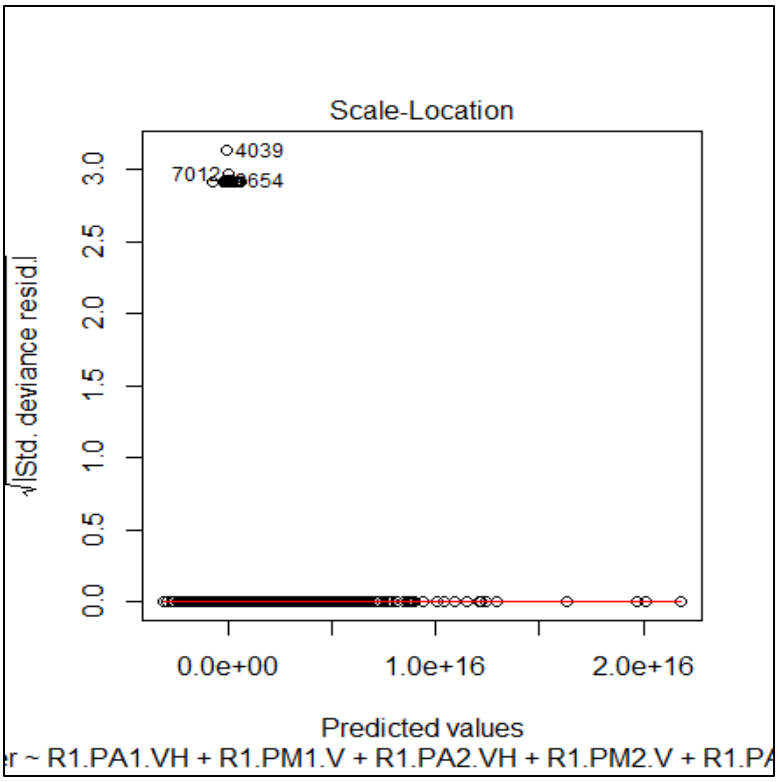
```
Start:  AIC=10462.4
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PM8.V + R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I +
    R1.PA11.IH + R1.PA12.IH + R1.PM12.I + R1.F + R1.DF + R1.PA.ZH +
    R1.S + R2.PA1.VH + R2.PM1.V + R2.PA2.VH + R2.PM2.V + R2.PA3.VH +
    R2.PM3.V + R2.PA4.IH + R2.PM4.I + R2.PA5.IH + R2.PM5.I +
    R2.PA6.IH + R2.PM6.I + R2.PA7.VH + R2.PM7.V + R2.PA8.VH +
    R2.PM8.V + R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I +
    R2.PA11.IH + R2.PM11.I + R2.PA12.IH + R2.PM12.I + R2.F +
    R2.DF + R2.PA.ZH + R2.S + R3.PA1.VH + R3.PM1.V + R3.PA2.VH +
    R3.PM2.V + R3.PA3.VH + R3.PM3.V + R3.PA4.IH + R3.PM4.I +
    R3.PA5.IH + R3.PM5.I + R3.PA6.IH + R3.PM6.I + R3.PA7.VH +
    R3.PA8.VH + R3.PM8.V + R3.PA9.VH + R3.PM9.V + R3.PA10.IH +
    R3.PM10.I + R3.PA11.IH + R3.PM11.I + R3.PA12.IH + R3.PM12.I +
    R3.F + R3.DF + R3.PA.ZH + R3.S + R4.PA1.VH + R4.PM1.V + R4.PA2.VH +
    R4.PM2.V + R4.PA3.VH + R4.PM3.V + R4.PA4.IH + R4.PM4.I +
    R4.PA5.IH + R4.PM5.I + R4.PA6.IH + R4.PM6.I + R4.PA7.VH +
    R4.PM7.V + R4.PA8.VH + R4.PM8.V + R4.PA9.VH + R4.PM9.V +
    R4.PA10.IH + R4.PM10.I + R4.PA11.IH + R4.PM11.I + R4.PA12.IH +
    R4.F + R4.DF + R4.PA.ZH + R4.S + R1.PA.Z + R2.PA.Z + R3.PA.Z +
    R4.PA.Z
```

```
Step:  AIC=4615.33
marker ~ R1.PA1.VH + R1.PM1.V + R1.PA2.VH + R1.PM2.V + R1.PA3.VH +
    R1.PM3.V + R1.PA4.IH + R1.PM4.I + R1.PA5.IH + R1.PM5.I +
    R1.PA6.IH + R1.PM6.I + R1.PA7.VH + R1.PM7.V + R1.PA8.VH +
    R1.PA9.VH + R1.PM9.V + R1.PA10.IH + R1.PM10.I + R1.PA11.IH +
    R1.PA12.IH + R1.PM12.I + R1.F + R1.DF + R1.PA.ZH + R1.S +
    R2.PA1.VH + R2.PM1.V + R2.PM2.V + R2.PA3.VH + R2.PM3.V +
    R2.PA4.IH + R2.PM4.I + R2.PA5.IH + R2.PM5.I + R2.PA6.IH +
    R2.PM6.I + R2.PA7.VH + R2.PM7.V + R2.PA8.VH + R2.PM8.V +
    R2.PA9.VH + R2.PM9.V + R2.PA10.IH + R2.PM10.I + R2.PA11.IH +
    R2.PM11.I + R2.PA12.IH + R2.PM12.I + R2.F + R2.DF + R2.PA.ZH +
    R2.S + R3.PA1.VH + R3.PM1.V + R3.PA2.VH + R3.PM2.V + R3.PA3.VH +
    R3.PM3.V + R3.PA4.IH + R3.PM4.I + R3.PA5.IH + R3.PM5.I +
    R3.PA6.IH + R3.PM6.I + R3.PA7.VH + R3.PA8.VH + R3.PM8.V +
    R3.PM9.V + R3.PA10.IH + R3.PM10.I + R3.PA11.IH + R3.PM11.I +
    R3.PA12.IH + R3.PM12.I + R3.F + R3.DF + R3.PA.ZH + R3.S +
    R4.PA1.VH + R4.PM1.V + R4.PA2.VH + R4.PM2.V + R4.PA3.VH +
    R4.PM3.V + R4.PA4.IH + R4.PM4.I + R4.PA5.IH + R4.PM5.I +
    R4.PA6.IH + R4.PM6.I + R4.PA7.VH + R4.PM7.V + R4.PA8.VH +
    R4.PM8.V + R4.PA9.VH + R4.PM9.V + R4.PA10.IH + R4.PM10.I +
    R4.PA11.IH + R4.PM11.I + R4.F + R4.DF + R4.PA.ZH + R1.PA.Z +
    R2.PA.Z + R3.PA.Z + R4.PA.Z

             Df Deviance     AIC
<none>            4397.3  4615.3
- R3.S        1   4469.4  4685.4
- R2.PA3.VH   1   4541.5  4757.5
- R3.PA2.VH   1   4757.8  4973.8
- R3.PA6.IH   1   4901.9  5117.9
- R4.PA.Z     1   4901.9  5117.9
- R1.PA4.IH   1   5046.1  5262.1
+ R2.PA2.VH   1   5118.2  5338.2
- R1.PM4.I    1   5190.3  5406.3
- R1.DF       1   5190.3  5406.3
- R1.PA3.VH   1   5262.4  5478.4
- R1.PM6.I    1   5334.5  5550.5
- R2.PM2.V    1   5334.5  5550.5
- R3.DF       1   5334.5  5550.5
- R1.PA1.VH   1   5550.7  5766.7
- R1.PA6.IH   1   5622.8  5838.8
- R3.PM8.V    1   5622.8  5838.8
- R4.DF       1   5622.8  5838.8
- R1.PM10.I   1   5694.9  5910.9
- R2.PM11.I   1   5694.9  5910.9
```

-	R3.PM5.I	1	5694.9	5910.9
-	R4.PA5.IH	1	5694.9	5910.9
-	R1.PM1.V	1	5767.0	5983.0
-	R3.PA3.VH	1	5767.0	5983.0
-	R4.PM9.V	1	5767.0	5983.0
-	R3.PM6.I	1	5839.1	6055.1
-	R4.PM10.I	1	5983.2	6199.2
-	R1.PA10.IH	1	6055.3	6271.3
-	R3.PA11.IH	1	6055.3	6271.3
-	R4.PM4.I	1	6055.3	6271.3
-	R1.F	1	6127.4	6343.4
-	R2.PA10.IH	1	6127.4	6343.4
-	R2.DF	1	6127.4	6343.4
-	R4.PA4.IH	1	6127.4	6343.4
-	R3.PM4.I	1	6199.5	6415.5
-	R1.PM5.I	1	6271.6	6487.6
-	R3.PA8.VH	1	6271.6	6487.6
+	R1.PM8.V	1	6271.6	6491.6
-	R2.PA9.VH	1	6343.7	6559.7
-	R3.PM11.I	1	6343.7	6559.7
-	R4.PA.ZH	1	6415.8	6631.8
-	R3.PM10.I	1	6487.9	6703.9
-	R4.PA6.IH	1	6487.9	6703.9
-	R2.PM3.V	1	6776.2	6992.2
-	R2.PA4.IH	1	6776.2	6992.2
-	R2.PA.ZH	1	6776.2	6992.2
-	R4.PM5.I	1	6776.2	6992.2
-	R2.PM12.I	1	6848.3	7064.3
-	R4.PA10.IH	1	6848.3	7064.3
-	R1.PM2.V	1	6920.4	7136.4
-	R2.PA6.IH	1	6920.4	7136.4
-	R2.PA8.VH	1	6920.4	7136.4
-	R2.F	1	6920.4	7136.4
-	R1.PM9.V	1	6992.5	7208.5
-	R1.PA5.IH	1	7064.6	7280.6
-	R1.PM3.V	1	7136.6	7352.6
-	R1.S	1	7208.7	7424.7
-	R4.PM6.I	1	7208.7	7424.7
-	R2.PM7.V	1	7280.8	7496.8
-	R3.PM2.V	1	7280.8	7496.8
-	R1.PA8.VH	1	7352.9	7568.9
-	R1.PA9.VH	1	7425.0	7641.0
-	R2.PM1.V	1	7425.0	7641.0
-	R4.PA3.VH	1	7425.0	7641.0

- R3.PA7.VH	1	7569.2	7785.2
- R4.PA2.VH	1	7785.4	8001.4
- R1.PM7.V	1	7929.6	8145.6
- R2.PM5.I	1	7929.6	8145.6
- R1.PA.Z	1	7929.6	8145.6
- R1.PA7.VH	1	8145.9	8361.9
- R2.PA5.IH	1	8218.0	8434.0
- R1.PA.ZH	1	8434.2	8650.2
- R3.PA5.IH	1	8434.2	8650.2
- R4.PM3.V	1	8434.2	8650.2
- R2.PA11.IH	1	8506.3	8722.3
- R4.PA7.VH	1	8506.3	8722.3
- R3.F	1	8650.5	8866.5
- R2.PA.Z	1	8650.5	8866.5
- R2.PM8.V	1	8722.6	8938.6
- R4.PA8.VH	1	8722.6	8938.6
- R2.PA1.VH	1	8794.7	9010.7
- R1.PM12.I	1	8866.7	9082.7
- R2.PA7.VH	1	8866.7	9082.7
- R3.PM1.V	1	8938.8	9154.8
- R1.PA2.VH	1	9010.9	9226.9
- R3.PA.ZH	1	9155.1	9371.1
- R4.F	1	9299.3	9515.3
- R4.PA1.VH	1	9371.3	9587.3
- R2.PM4.I	1	9587.6	9803.6
- R4.PM8.V	1	9731.8	9947.8
- R4.PA9.VH	1	9731.8	9947.8
- R2.PM6.I	1	9803.9	10019.9
- R3.PA1.VH	1	10020.1	10236.1
- R3.PA10.IH	1	10092.2	10308.2
- R1.PA11.IH	1	10092.2	10308.2
- R4.PA11.IH	1	10164.3	10380.3
- R3.PM3.V	1	10236.4	10452.4
- R4.PM1.V	1	10308.5	10524.5
- R3.PM9.V	1	10380.6	10596.6
- R3.PA12.IH	1	10380.6	10596.6
- R2.PM10.I	1	10452.7	10668.7

- R2.S	1	10452.7	10668.7
- R1.PA12.IH	1	10957.3	11173.3
- R2.PM9.V	1	11101.4	11317.4
- R4.PM7.V	1	11101.4	11317.4
- R4.PM11.I	1	11389.8	11605.8
- R2.PA12.IH	1	11678.1	11894.1
- R4.PM2.V	1	11894.4	12110.4
+ R4.PA12.IH	1	11894.4	12114.4
- R3.PA4.IH	1	11966.5	12182.5
- R3.PM12.I	1	12254.8	12470.8
+ R3.PA9.VH	1	13984.9	14204.9
- R3.PA.Z	1	14417.5	14633.5

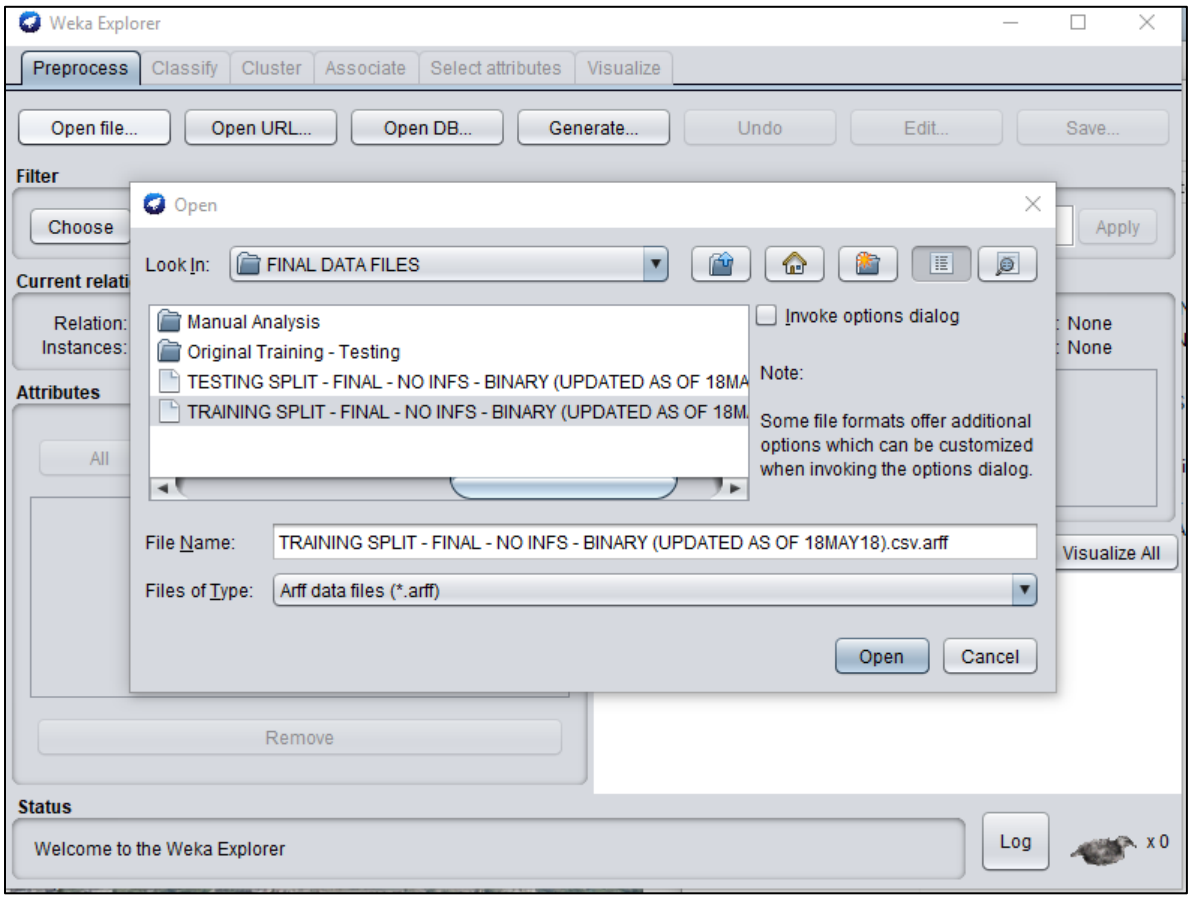


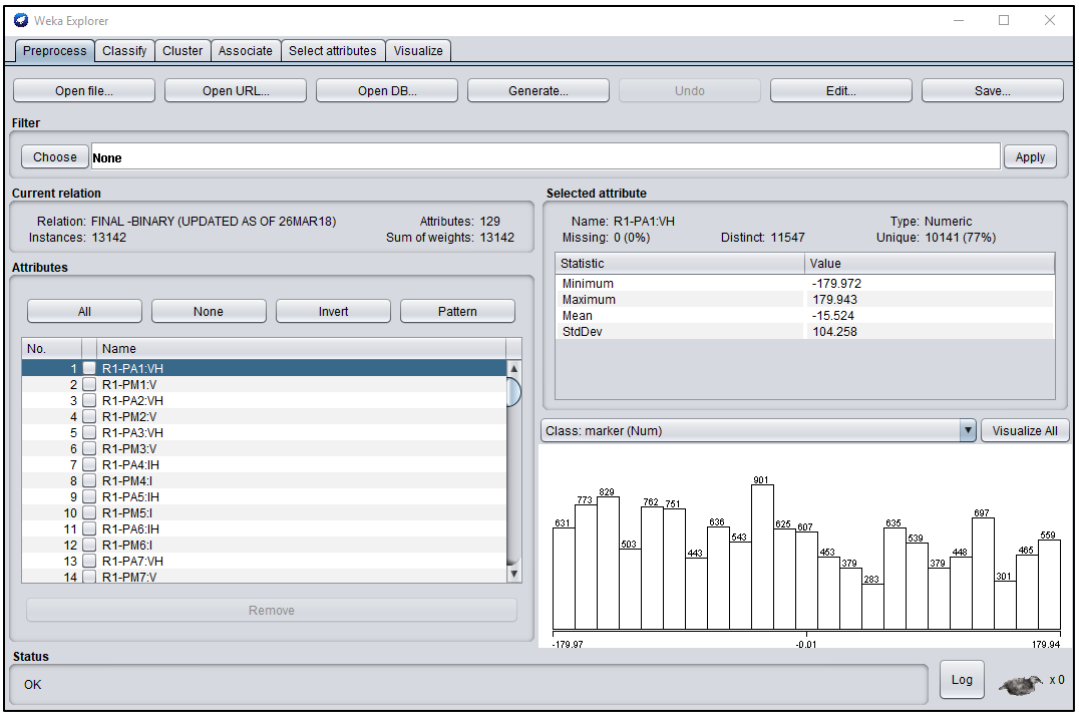

```
easystdris4 <- easystdstepwisemodel4$residuals
summary(easystdris4)


sum(filter(std.easy, easystdris4>2)$marker==1)
sum(filter(std.easy, easystdris4>2)$marker==0)

##OUTPUT: 26 of the residuals over the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 0 of the residuals over the 2.0 mark from std easy dataset are normal operations

sum(filter(std.easy, easystdris4<2)$marker==1)
sum(filter(std.easy, easystdris4<2)$marker==0)

##OUTPUT: 5242 of the residuals less than the 2.0 mark (anomalies) from std easy dataset are attacks
##OUPUT: 4405 of the residuals less than the 2.0 mark from std easy dataset are normal operations
```

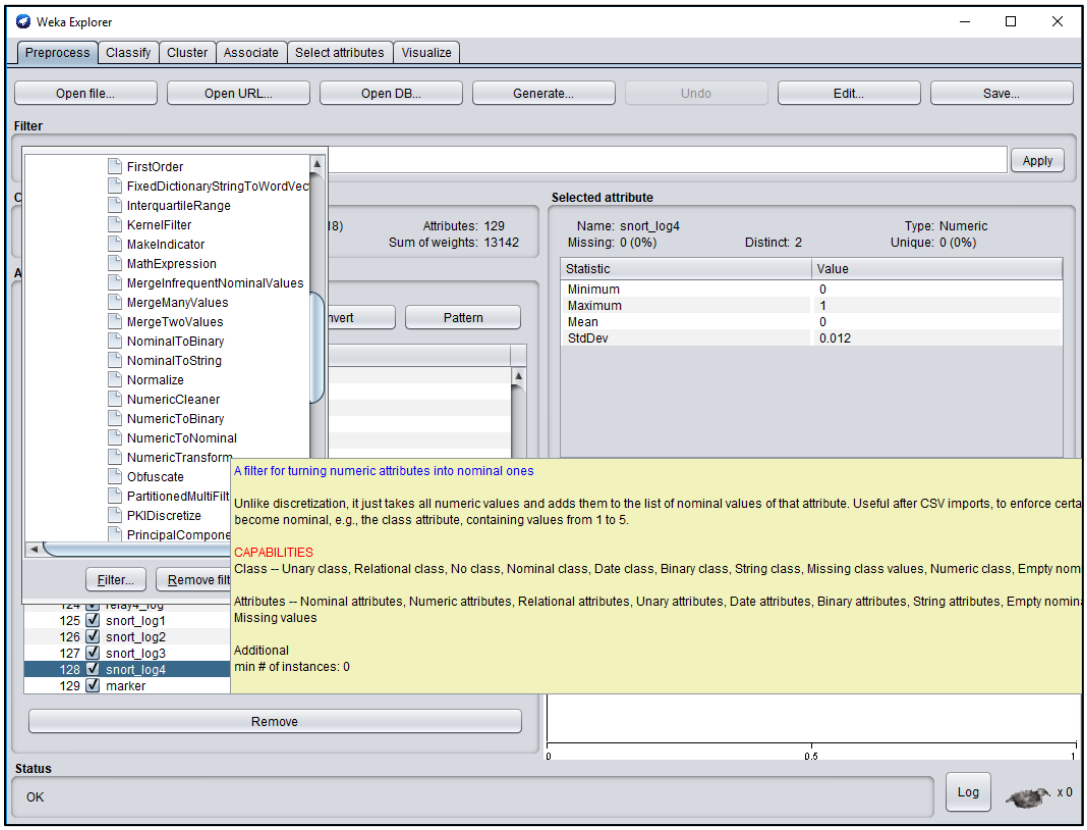
APPENDIX H. LOADING AND PREPROCESSING DATA IN WEKA

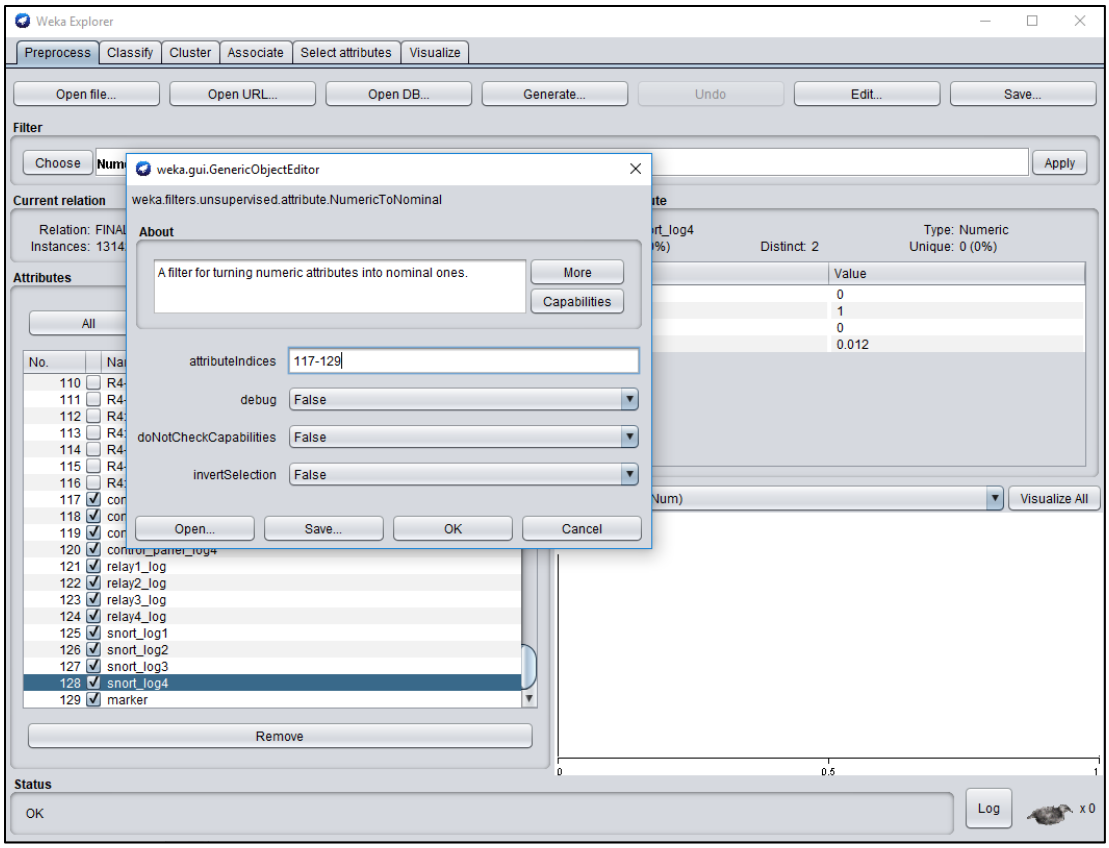
After opening WEKA software, the Explorer option was chosen (see first image below). Utilizing the “Open file” button in the upper right hand corner in the Preprocess tab in the GUI, the training dataset was selected and opened (TRAINING SPLIT – FINAL – NO INFS – BINARY (UPDATED AS OF 18MAY18).csv,arff; see second image below). After loading the dataset, variables (referred to as attributes in WEKA) are displayed with the minimum values, maximum values, mean, and standard deviation in the GUI (see third image below; the first variable R1-PA1-VH is displayed). Upon observing the data, it was noticed that multiple variables were incorrectly classified by data type. Specifically, that nominal/binary values were being read in to WEKA as numeric values (variables 117-129, all log files and the marker). To address this, the variables whose data types were incorrect were selected and the filter NumerictoNominal was applied (see fourth image below). The attributes/indices were then modified to incorporate the applicable variables by right clicking in the space and selecting “show properties” and then “OK” (see fifth image below). After closing the properties box, the “Apply” button was selected in order to change the data types of the selected variables from numeric to nominal. The output of a selected variable was automatically colored after this step (blue for normal operations, red for an attack) and the data type in the selected attribute panel reflects a nominal data type (see sixth image below). This concludes the steps taken to preprocess the data in WEKA prior to classification of the dataset.

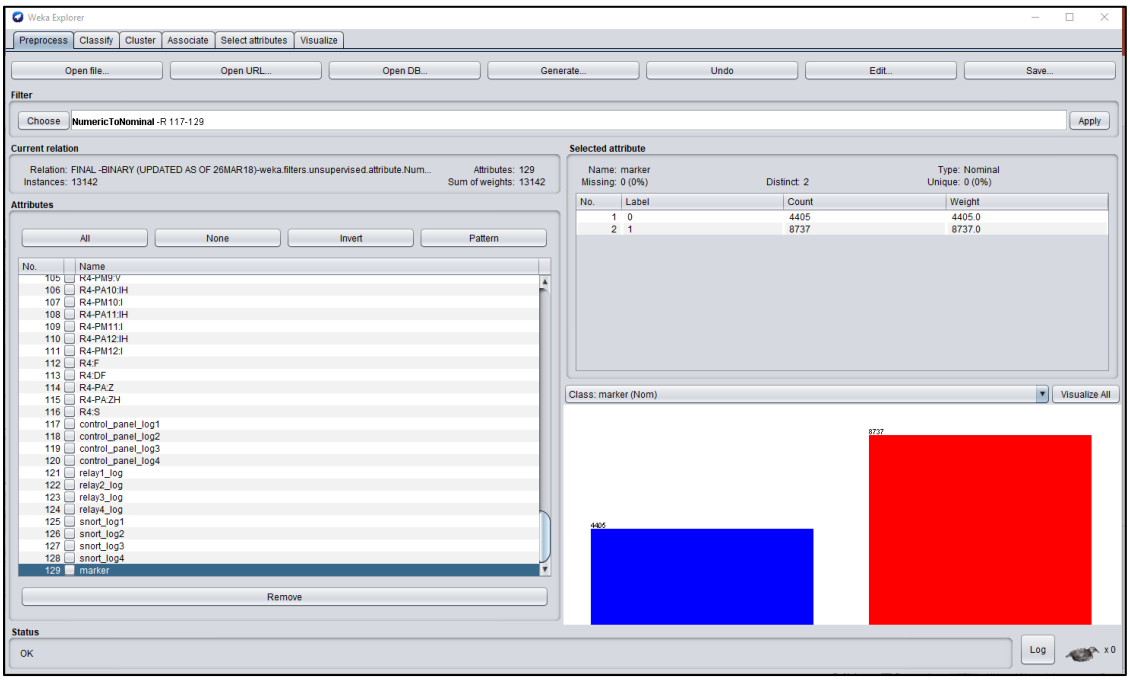

Weka GUI Chooser
Program Visualization Tools Help
WEKA
The University of Waikato
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI
Waikato Environment for Knowledge Analysis
Version 3.8.1
(c) 1999 - 2016
The University of Waikato
Hamilton, New Zealand


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Apply
Current relati
Relation:
Instances:
Attributes
All
Remove
Visualize All
Status
Welcome to the Weka Explorer
Log
x 0
Open
Look In:
FINAL DATA FILES
Manual Analysis
Original Training - Testing
TESTING SPLIT - FINAL - NO INFS - BINARY (UPDATED AS OF 18MA
TRAINING SPLIT - FINAL - NO INFS - BINARY (UPDATED AS OF 18M
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
TRAINING SPLIT - FINAL - NO INFS - BINARY (UPDATED AS OF 18MAY18).csv.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: FINAL -BINARY (UPDATED AS OF 26MAR18)
Instances: 13142
Attributes: 129
Sum of weights: 13142
Attributes
All
None
Invert
Pattern
No.
Name
1 R1-PA1:VH
2 R1-PM1:V
3 R1-PA2:VH
4 R1-PM2:V
5 R1-PA3:VH
6 R1-PM3:V
7 R1-PA4:IH
8 R1-PM4:I
9 R1-PA5:IH
10 R1-PM5:I
11 R1-PA6:IH
12 R1-PM6:I
13 R1-PA7:VH
14 R1-PM7:V
Remove
Selected attribute
Name: R1-PA1:VH
Missing: 0 (0%)
Distinct: 11547
Type: Numeric
Unique: 10141 (77%)
Statistic
Value
Minimum
-179.972
Maximum
179.943
Mean
-15.524
StdDev
104.258
Class: marker (Num)
Visualize All
631
773
829
503
762
751
443
636
543
901
625
607
453
379
283
635
539
379
448
697
301
465
559
-179.97
-0.01
179.94
Status
OK
Log
x 0

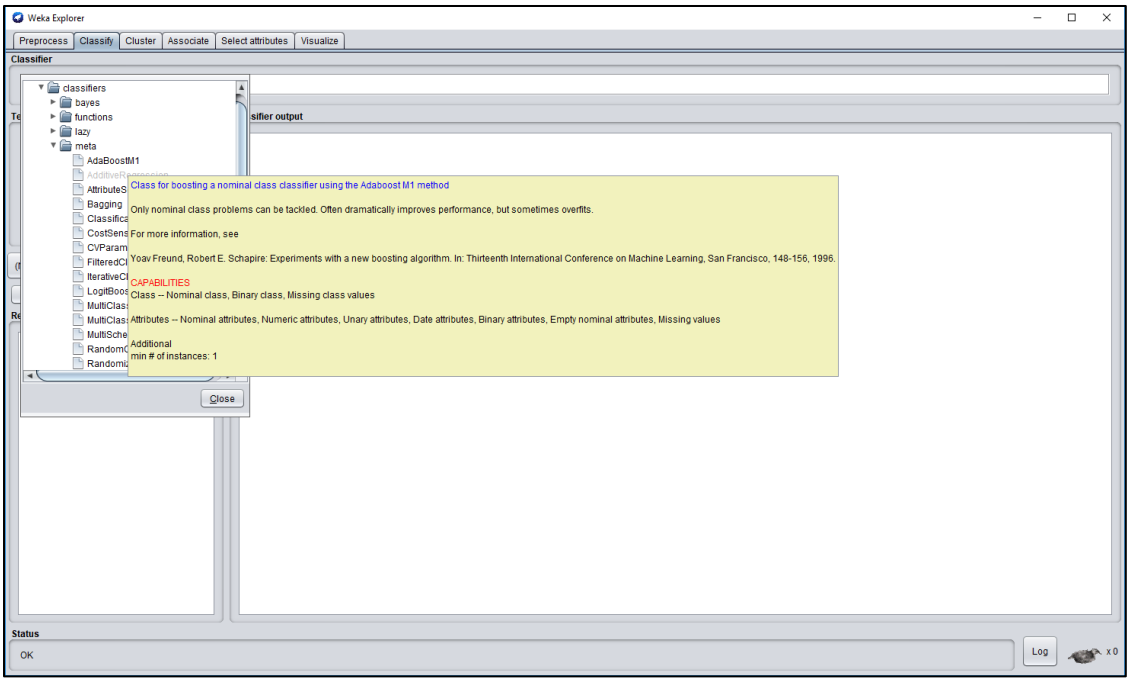

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Apply
FirstOrder
FixedDictionaryStringToWordVec
InterquartileRange
KernelFilter
MakeIndicator
MathExpression
MergeInfrequentNominalValues
MergeManyValues
MergeTwoValues
NominalToBinary
NominalToString
Normalize
NumericCleaner
NumericToBinary
NumericToNominal
NumericTransform
Obfuscate
PartitionedMultiFilt
PKIDiscretize
PrincipalCompone
Filter...
Remove filt
Attributes: 129
Sum of weights: 13142
Pattern
Selected attribute
Name: snort_log4
Missing: 0 (0%)
Distinct: 2
Type: Numeric
Unique: 0 (0%)
Statistic
Value
Minimum
0
Maximum
1
Mean
0
StdDev
0.012
A filter for turning numeric attributes into nominal ones
Unlike discretization, it just takes all numeric values and adds them to the list of nominal values of that attribute. Useful after CSV imports, to enforce certa
become nominal, e.g., the class attribute, containing values from 1 to 5.
CAPABILITIES
Class -- Unary class, Relational class, No class, Nominal class, Date class, Binary class, String class, Missing class values, Numeric class, Empty nom
Attributes -- Nominal attributes, Numeric attributes, Relational attributes, Unary attributes, Date attributes, Binary attributes, String attributes, Empty nomina
Missing values
Additional
min # of instances: 0
125 snort_log1
126 snort_log2
127 snort_log3
128 snort_log4
129 marker
Remove
0
0.5
1
Status
OK
Log
x 0

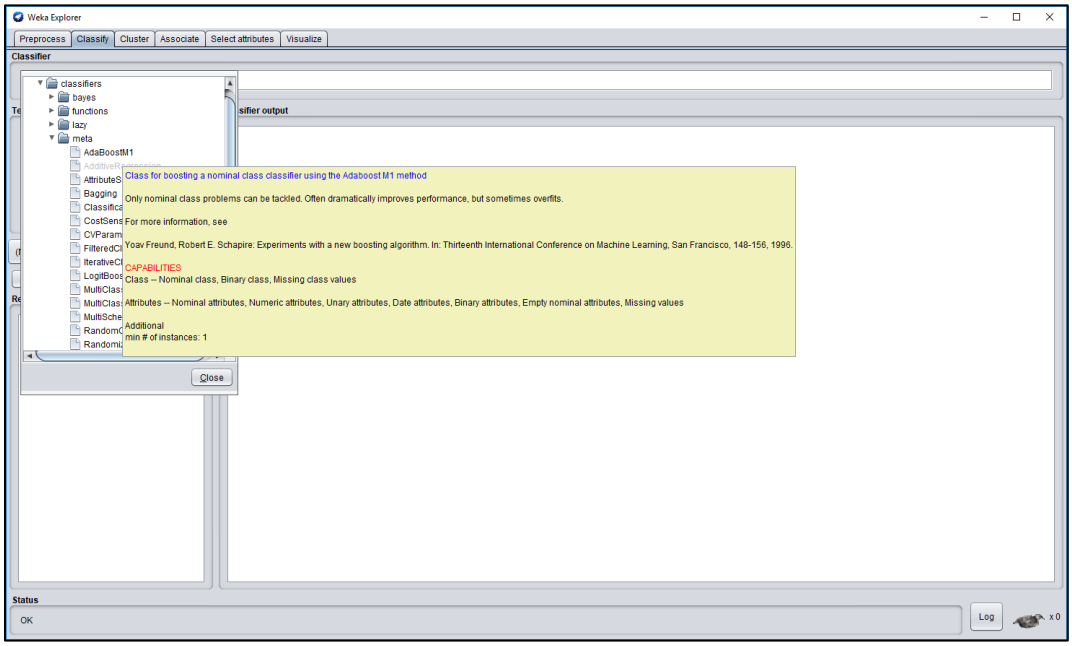

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Apply
Current relation
Attributes
All
Remove
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.NumericToNominal
About
A filter for turning numeric attributes into nominal ones.
More
Capabilities
attributeIndices
117-129
debug
False
doNotCheckCapabilities
False
invertSelection
False
Open...
Save...
OK
Cancel
Type: Numeric
Unique: 0 (0%)
Distinct: 2
Value
0
1
0
0.012
Visualize All
110 R4-
111 R4-
112 R4-
113 R4-
114 R4-
115 R4-
116 R4-
117 cor
118 cor
119 cor
120 control_panel_log4
121 relay1_log
122 relay2_log
123 relay3_log
124 relay4_log
125 snort_log1
126 snort_log2
127 snort_log3
128 snort_log4
129 marker

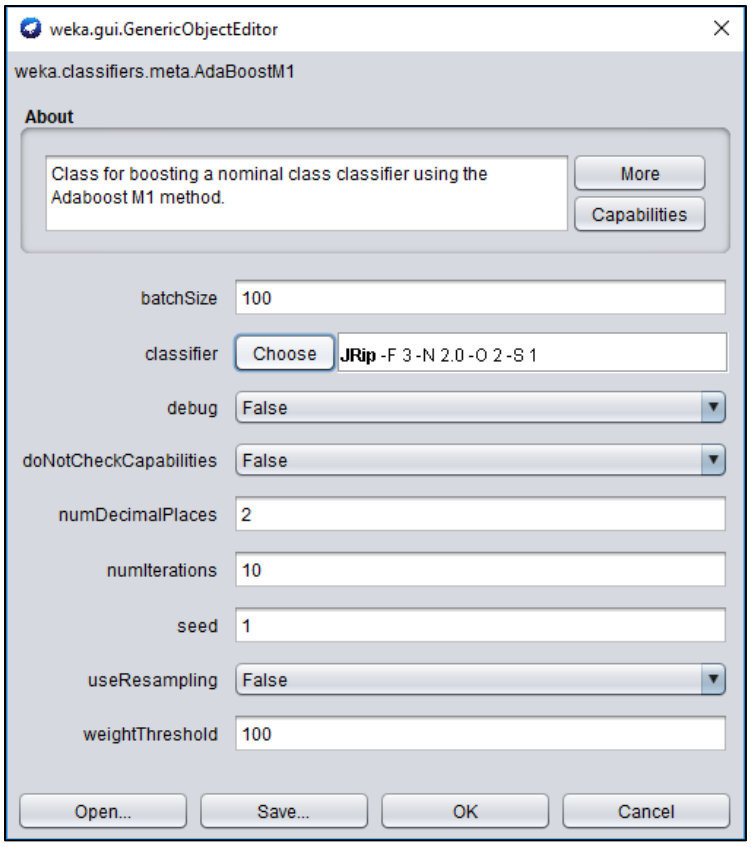

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
NumericToNominal -R 117-129
Apply
Current relation
Relation: FINAL -BINARY (UPDATED AS OF 26MAR18)-weka.filters.unsupervised.attribute.Num...
Instances: 13142
Attributes: 129
Sum of weights: 13142
Attributes
All
None
Invert
Pattern
Remove
Status
OK
Log
x 0
Selected attribute
Name: marker
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No. Label Count Weight
1 0 4405 4405.0
2 1 8737 8737.0
Class: marker (Nom)
Visualize All
8737
4405
105 R4-PM9:V
106 R4-PA10:IH
107 R4-PM10:I
108 R4-PA11:IH
109 R4-PM11:I
110 R4-PA12:IH
111 R4-PM12:I
112 R4:F
113 R4:DF
114 R4-PA:Z
115 R4-PA:ZH
116 R4:S
117 control_panel_log1
118 control_panel_log2
119 control_panel_log3
120 control_panel_log4
121 relay1_log
122 relay2_log
123 relay3_log
124 relay4_log
125 snort_log1
126 snort_log2
127 snort_log3
128 snort_log4
129 marker

APPENDIX I. DATASET CLASSIFICATION IN WEKA

After preprocessing the dataset in WEKA, the “Classify” tab was selected, and under the “Classifier” panel various classifier were selected to compare performance (the image below only shows AdaboostM1+JRipper). AdaboostM1+JRipper was first selected using the “Choose” button under the “meta” classifiers folder, and the properties were modified to utilize this classifier in conjunction with JRipper. This was done via right clicking in the space next to AdaboostM1 next to the “Choose” button and selecting “Show Properties” (see second image below). The classifier JRipper was then selected by clicking the “Choose” button next to the classifier field, and then selecting JRipper from the “rules” classifier folder and then clicking “OK” (see third image below). The classification was then initiated via clicking the “Start” button in the “Test options” panel.

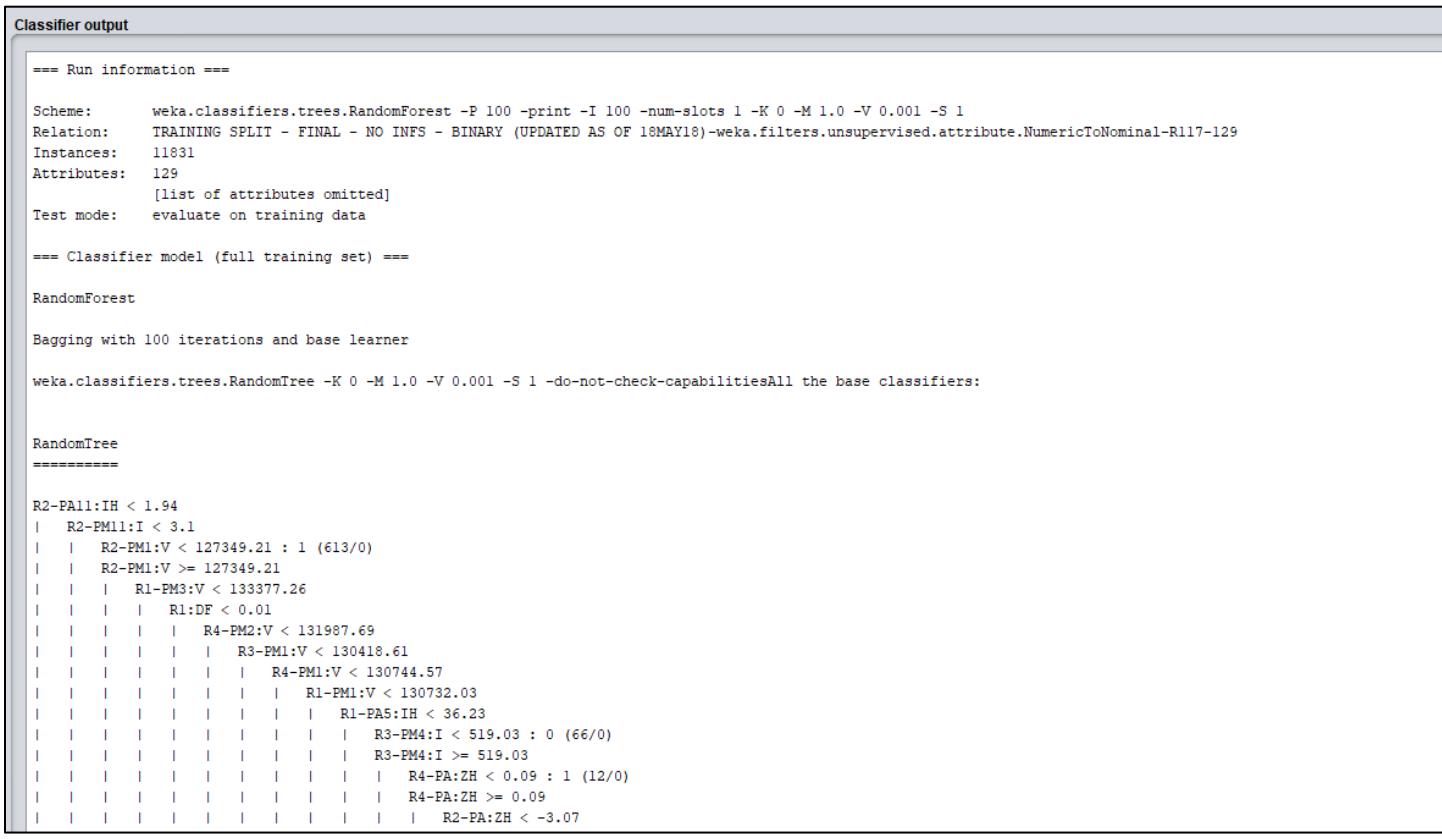

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
classifiers
bayes
functions
lazy
meta
AdaBoostM1
Class for boosting a nominal class classifier using the Adaboost M1 method
Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes overfits.
For more information, see
Yoav Freund, Robert E. Schapire: Experiments with a new boosting algorithm. In: Thirteenth International Conference on Machine Learning, San Francisco, 148-156, 1996.
CAPABILITIES
Class -- Nominal class, Binary class, Missing class values
Attributes -- Nominal attributes, Numeric attributes, Unary attributes, Date attributes, Binary attributes, Empty nominal attributes, Missing values
Additional
min # of instances: 1
Close
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
classifiers
bayes
functions
lazy
meta
AdaBoostM1
Class for boosting a nominal class classifier using the Adaboost M1 method
Only nominal class problems can be tackled. Often dramatically improves performance, but sometimes overfits.
For more information, see
Yoav Freund, Robert E. Schapire: Experiments with a new boosting algorithm. In: Thirteenth International Conference on Machine Learning, San Francisco, 148-156, 1996.
CAPABILITIES
Class -- Nominal class, Binary class, Missing class values
Attributes -- Nominal attributes, Numeric attributes, Unary attributes, Date attributes, Binary attributes, Empty nominal attributes, Missing values
Additional
min # of instances: 1
Close
Status
OK
Log
x 0


weka.gui.GenericObjectEditor
weka.classifiers.meta.AdaBoostM1
About
Class for boosting a nominal class classifier using the Adaboost M1 method.
More
Capabilities
batchSize
100
classifier
Choose
JRip -F 3 -N 2.0 -O 2 -S 1
debug
False
doNotCheckCapabilities
False
numDecimalPlaces
2
numIterations
10
seed
1
useResampling
False
weightThreshold
100
Open...
Save...
OK
Cancel

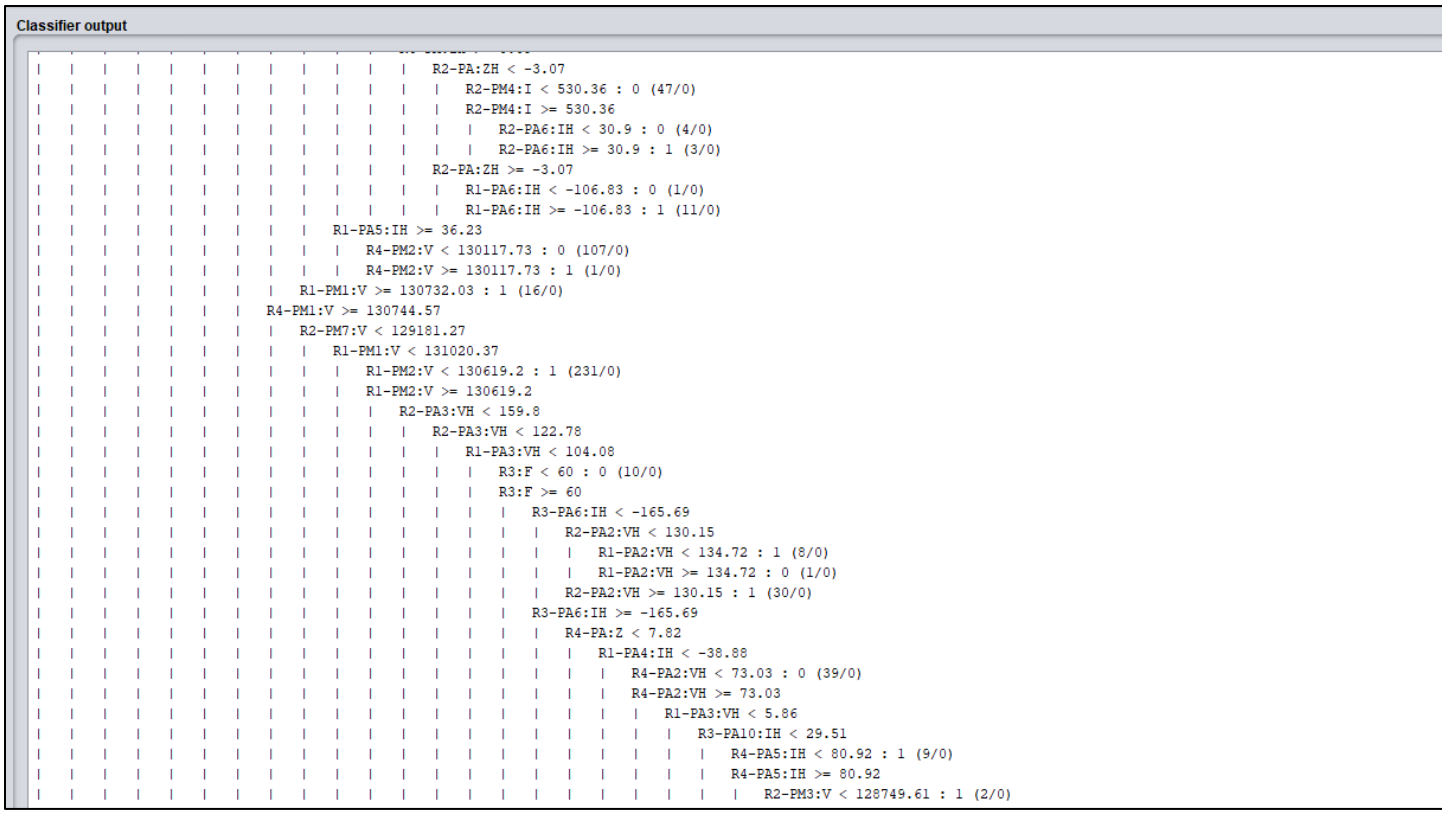
APPENDIX J. RANDOMFOREST ML ALGORITHM OUTPUTS

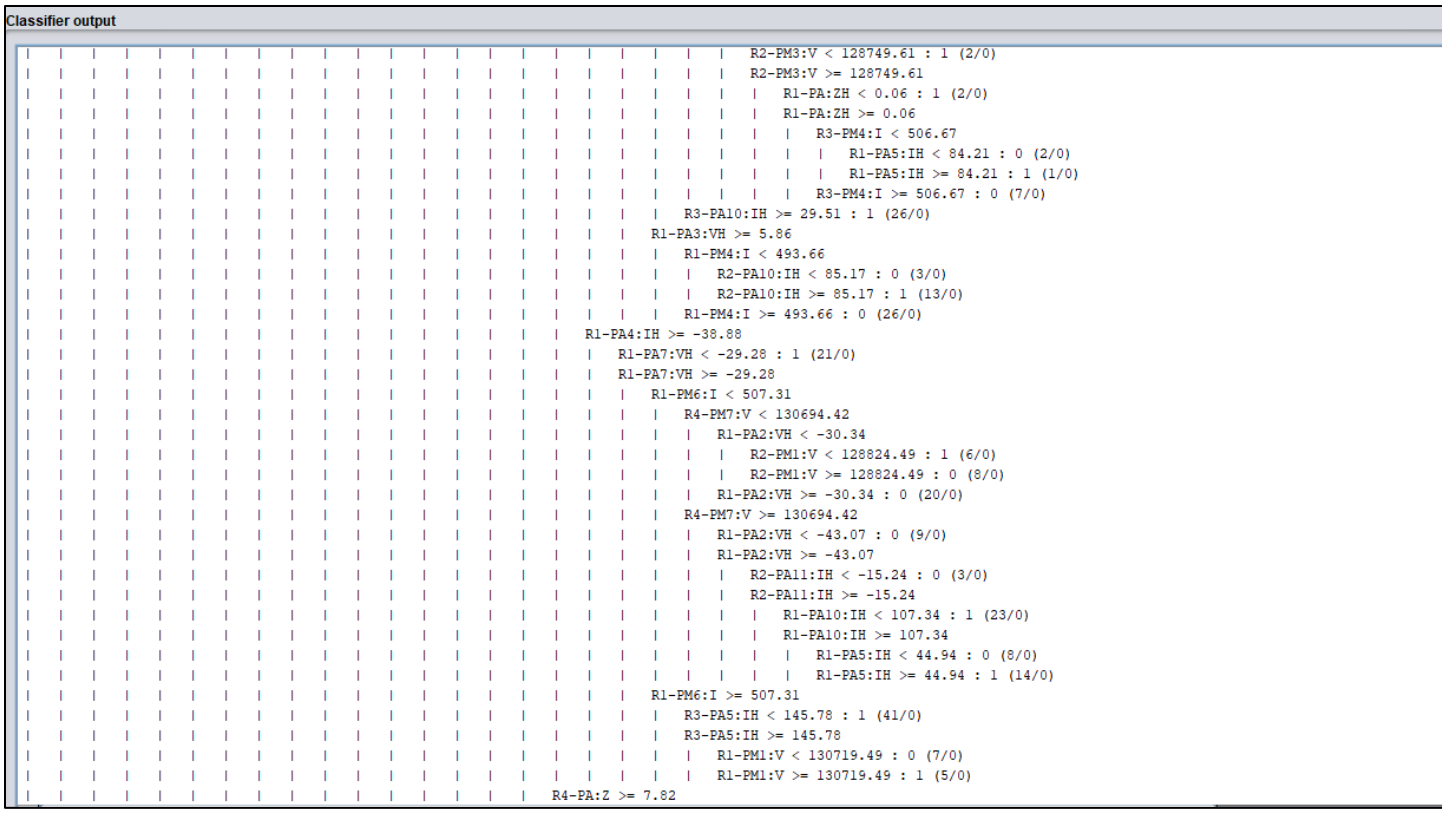
The WEKA output below includes the first Random Tree from the RandomForest algorithm, attribute importance, and summary statistics. Due to the large size of the output consisting of 100 Random Trees, only the first will be displayed for brevity.

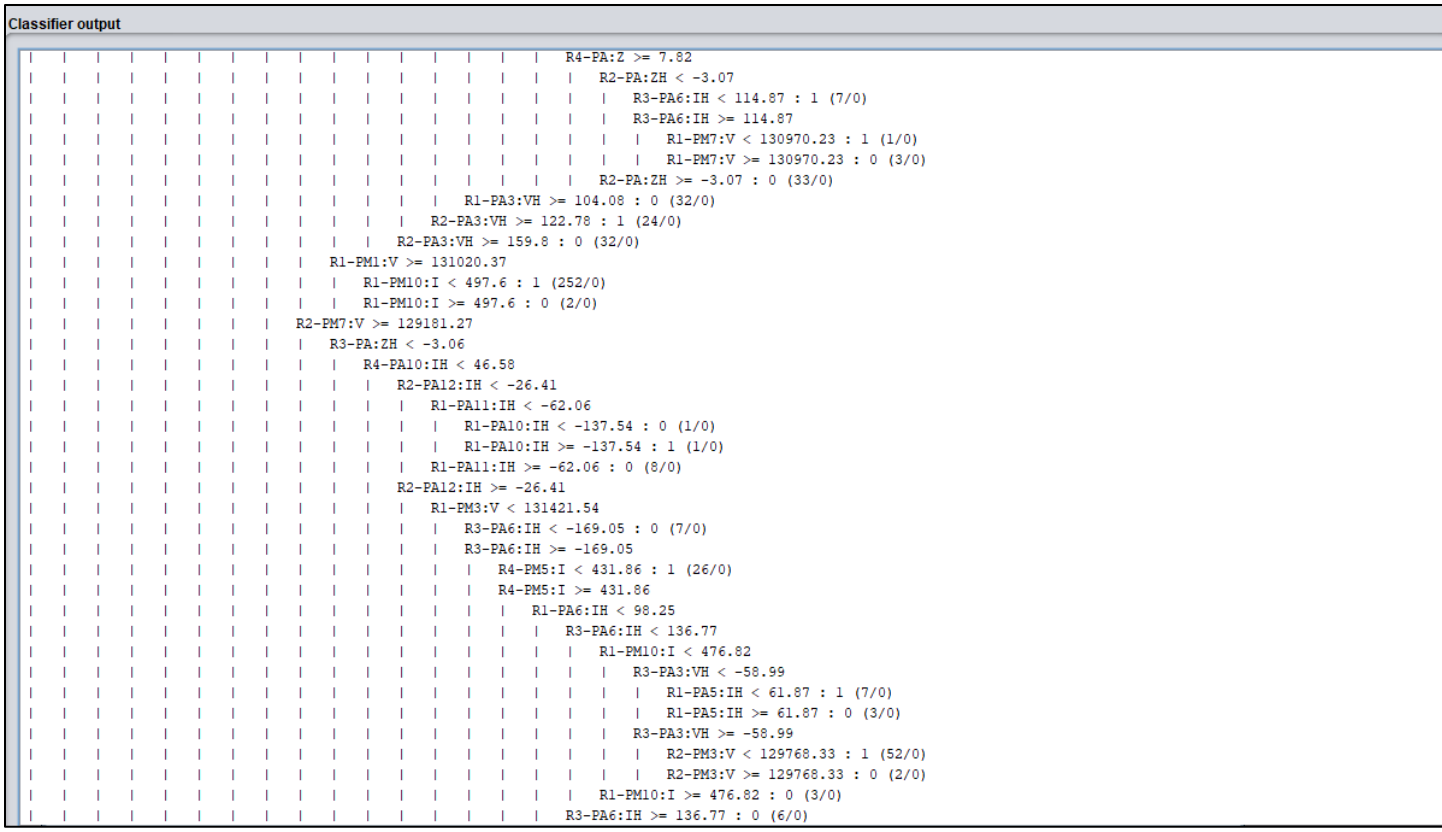


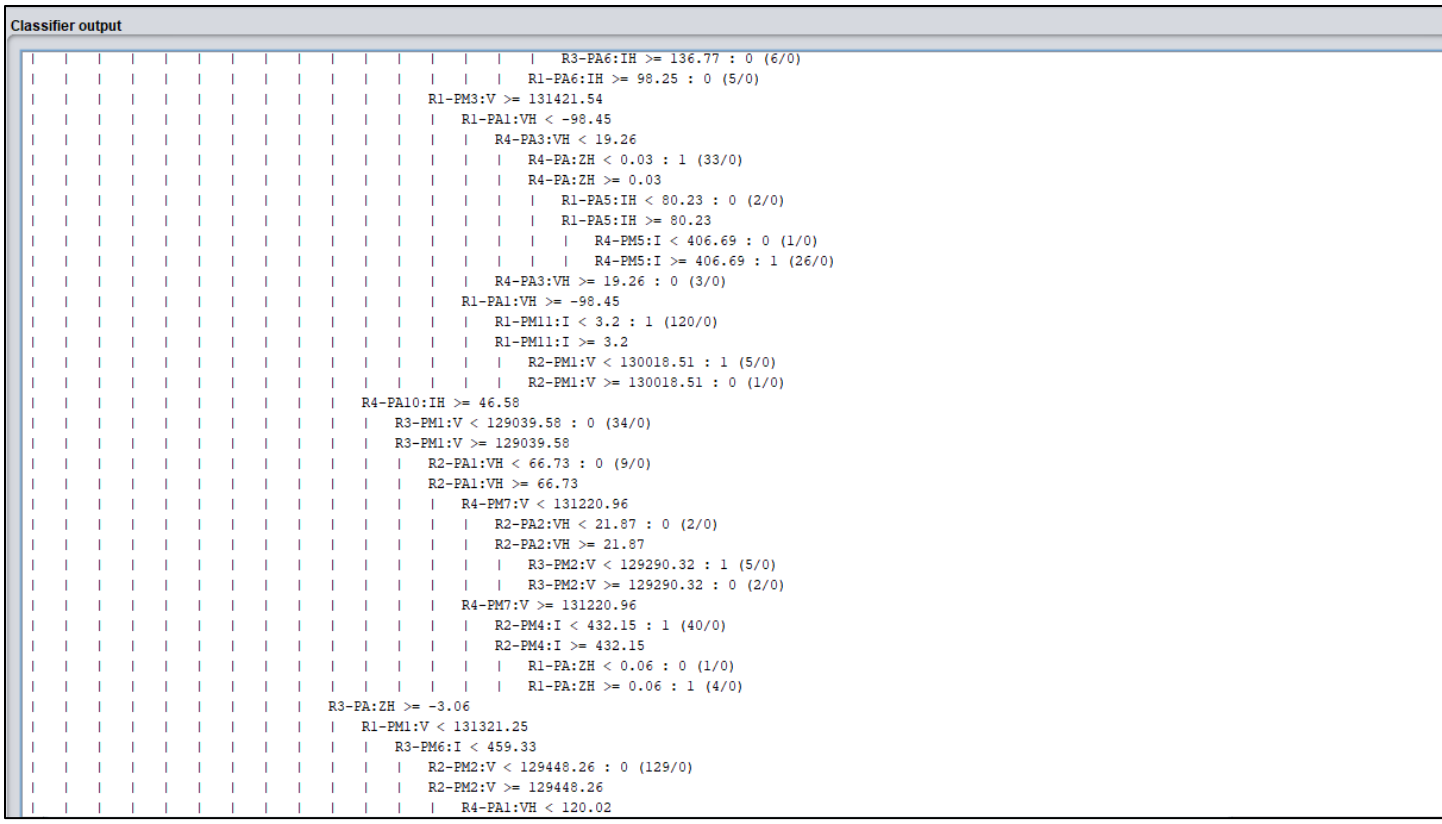
Classifier output

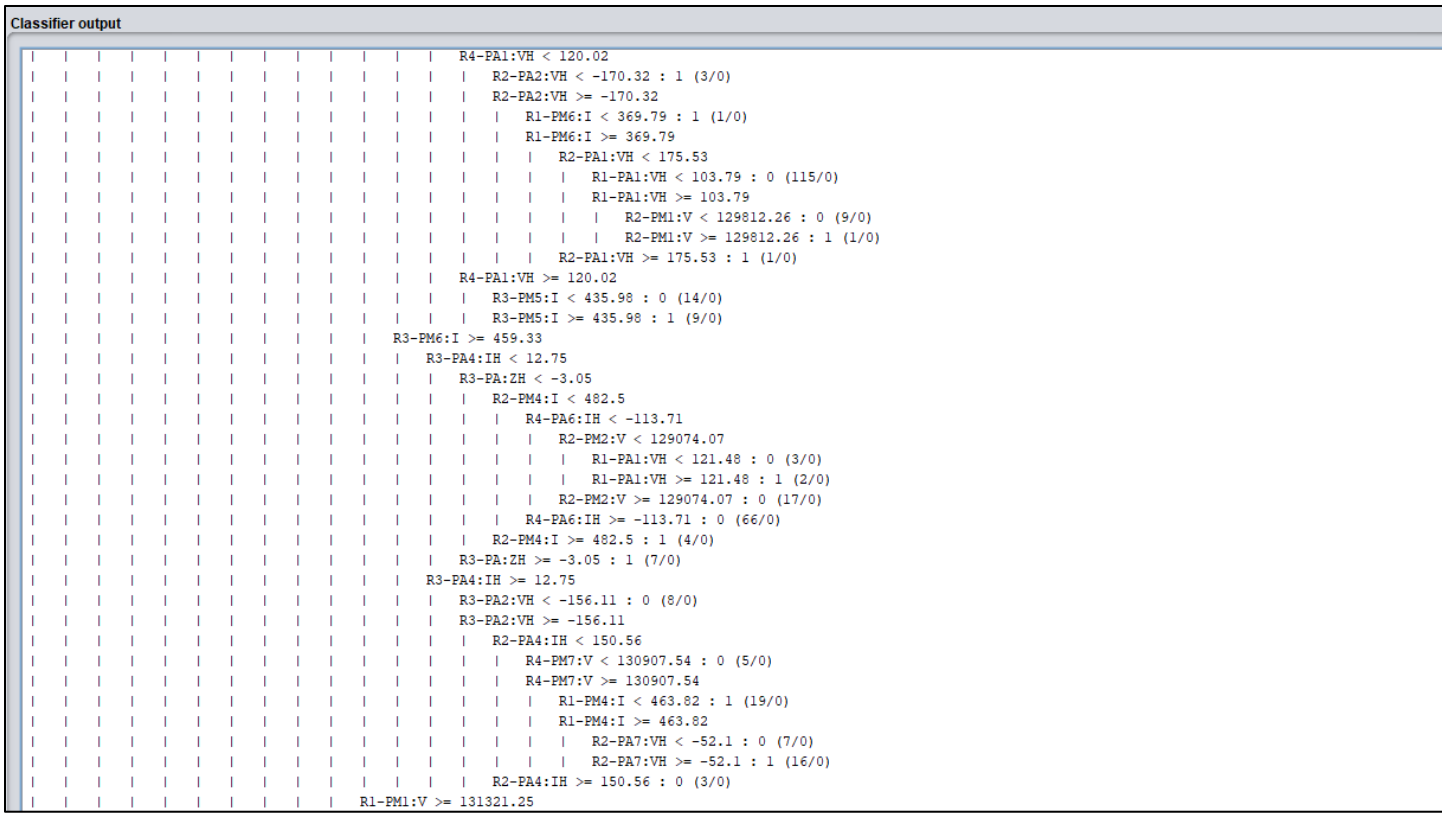
```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomForest -P 100 -print -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     TRAINING SPLIT - FINAL - NO INFS - BINARY (UPDATED AS OF 18MAY18)-weka.filters.unsupervised.attribute.NumericToNominal-R117-129
Instances:    11831
Attributes:   129
              [list of attributes omitted]
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilitiesAll the base classifiers:


RandomTree
==========

R2-PA11:IH < 1.94
|   R2-PM11:I < 3.1
|   |   R2-PM1:V < 127349.21 : 1 (613/0)
|   |   R2-PM1:V >= 127349.21
|   |   |   R1-PM3:V < 133377.26
|   |   |   |   R1:DF < 0.01
|   |   |   |   |   R4-PM2:V < 131987.69
|   |   |   |   |   |   R3-PM1:V < 130418.61
|   |   |   |   |   |   |   R4-PM1:V < 130744.57
|   |   |   |   |   |   |   |   R1-PM1:V < 130732.03
|   |   |   |   |   |   |   |   |   R1-PA5:IH < 36.23
|   |   |   |   |   |   |   |   |   |   R3-PM4:I < 519.03 : 0 (66/0)
|   |   |   |   |   |   |   |   |   |   R3-PM4:I >= 519.03
|   |   |   |   |   |   |   |   |   |   |   R4-PA:ZH < 0.09 : 1 (12/0)
|   |   |   |   |   |   |   |   |   |   |   R4-PA:ZH >= 0.09
|   |   |   |   |   |   |   |   |   |   |   |   R2-PA:ZH < -3.07
```

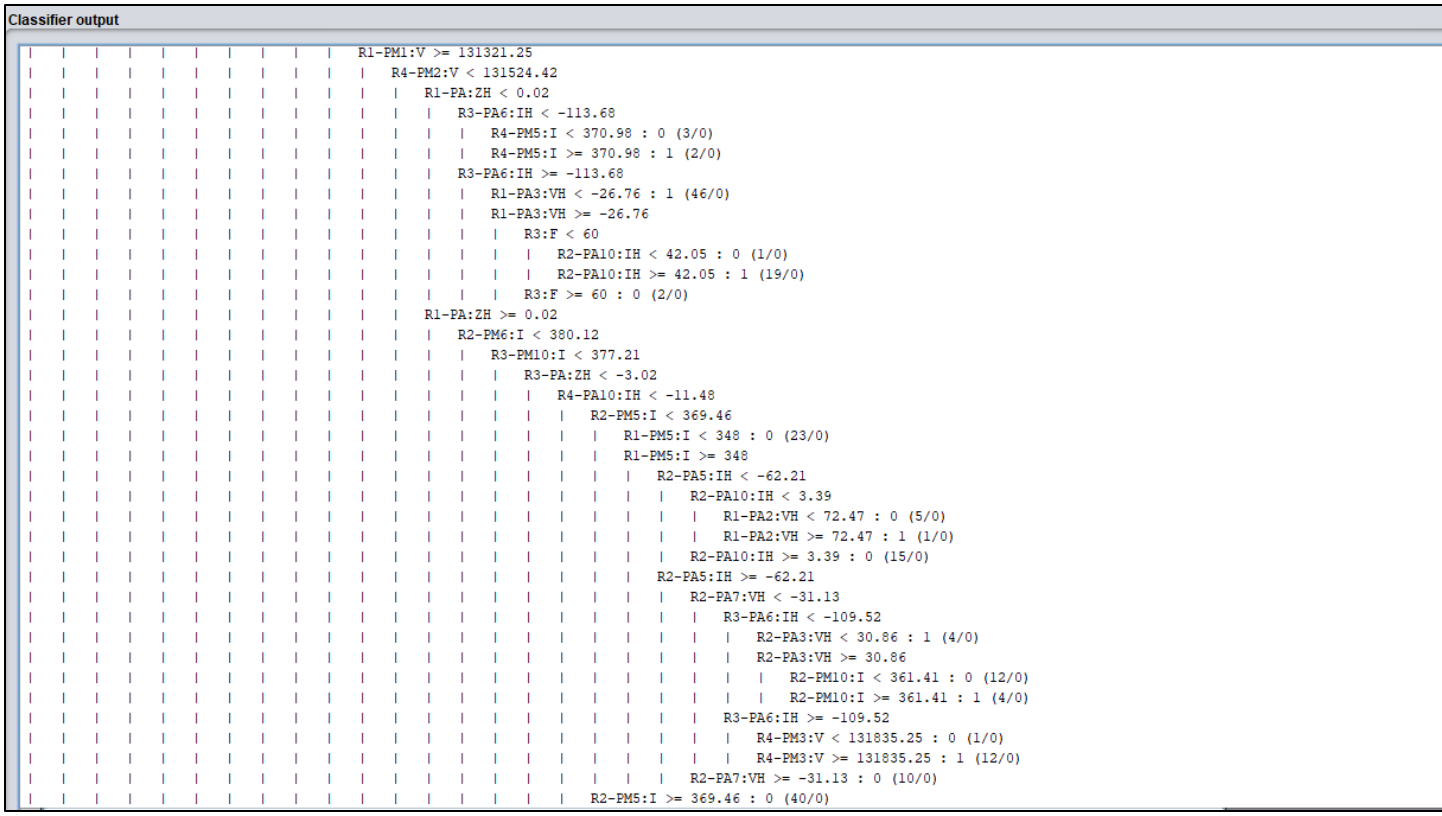

Classifier output
| | | | | | | | | | | | | R2-PA:ZH < -3.07
| | | | | | | | | | | | | | R2-PM4:I < 530.36 : 0 (47/0)
| | | | | | | | | | | | | | R2-PM4:I >= 530.36
| | | | | | | | | | | | | | | R2-PA6:IH < 30.9 : 0 (4/0)
| | | | | | | | | | | | | | | R2-PA6:IH >= 30.9 : 1 (3/0)
| | | | | | | | | | | | | R2-PA:ZH >= -3.07
| | | | | | | | | | | | | | R1-PA6:IH < -106.83 : 0 (1/0)
| | | | | | | | | | | | | | R1-PA6:IH >= -106.83 : 1 (11/0)
| | | | | | | | | | | R1-PA5:IH >= 36.23
| | | | | | | | | | | | R4-PM2:V < 130117.73 : 0 (107/0)
| | | | | | | | | | | | R4-PM2:V >= 130117.73 : 1 (1/0)
| | | | | | | | | | R1-PM1:V >= 130732.03 : 1 (16/0)
| | | | | | | | | R4-PM1:V >= 130744.57
| | | | | | | | | | R2-PM7:V < 129181.27
| | | | | | | | | | | R1-PM1:V < 131020.37
| | | | | | | | | | | | R1-PM2:V < 130619.2 : 1 (231/0)
| | | | | | | | | | | | R1-PM2:V >= 130619.2
| | | | | | | | | | | | | R2-PA3:VH < 159.8
| | | | | | | | | | | | | | R2-PA3:VH < 122.78
| | | | | | | | | | | | | | | R1-PA3:VH < 104.08
| | | | | | | | | | | | | | | | R3:F < 60 : 0 (10/0)
| | | | | | | | | | | | | | | | R3:F >= 60
| | | | | | | | | | | | | | | | | R3-PA6:IH < -165.69
| | | | | | | | | | | | | | | | | | R2-PA2:VH < 130.15
| | | | | | | | | | | | | | | | | | | R1-PA2:VH < 134.72 : 1 (8/0)
| | | | | | | | | | | | | | | | | | | R1-PA2:VH >= 134.72 : 0 (1/0)
| | | | | | | | | | | | | | | | | | R2-PA2:VH >= 130.15 : 1 (30/0)
| | | | | | | | | | | | | | | | | R3-PA6:IH >= -165.69
| | | | | | | | | | | | | | | | | | R4-PA:Z < 7.82
| | | | | | | | | | | | | | | | | | | R1-PA4:IH < -38.88
| | | | | | | | | | | | | | | | | | | | R4-PA2:VH < 73.03 : 0 (39/0)
| | | | | | | | | | | | | | | | | | | | R4-PA2:VH >= 73.03
| | | | | | | | | | | | | | | | | | | | | R1-PA3:VH < 5.86
| | | | | | | | | | | | | | | | | | | | | | R3-PA10:IH < 29.51
| | | | | | | | | | | | | | | | | | | | | | | R4-PA5:IH < 80.92 : 1 (9/0)
| | | | | | | | | | | | | | | | | | | | | | | R4-PA5:IH >= 80.92
| | | | | | | | | | | | | | | | | | | | | | | | R2-PM3:V < 128749.61 : 1 (2/0)

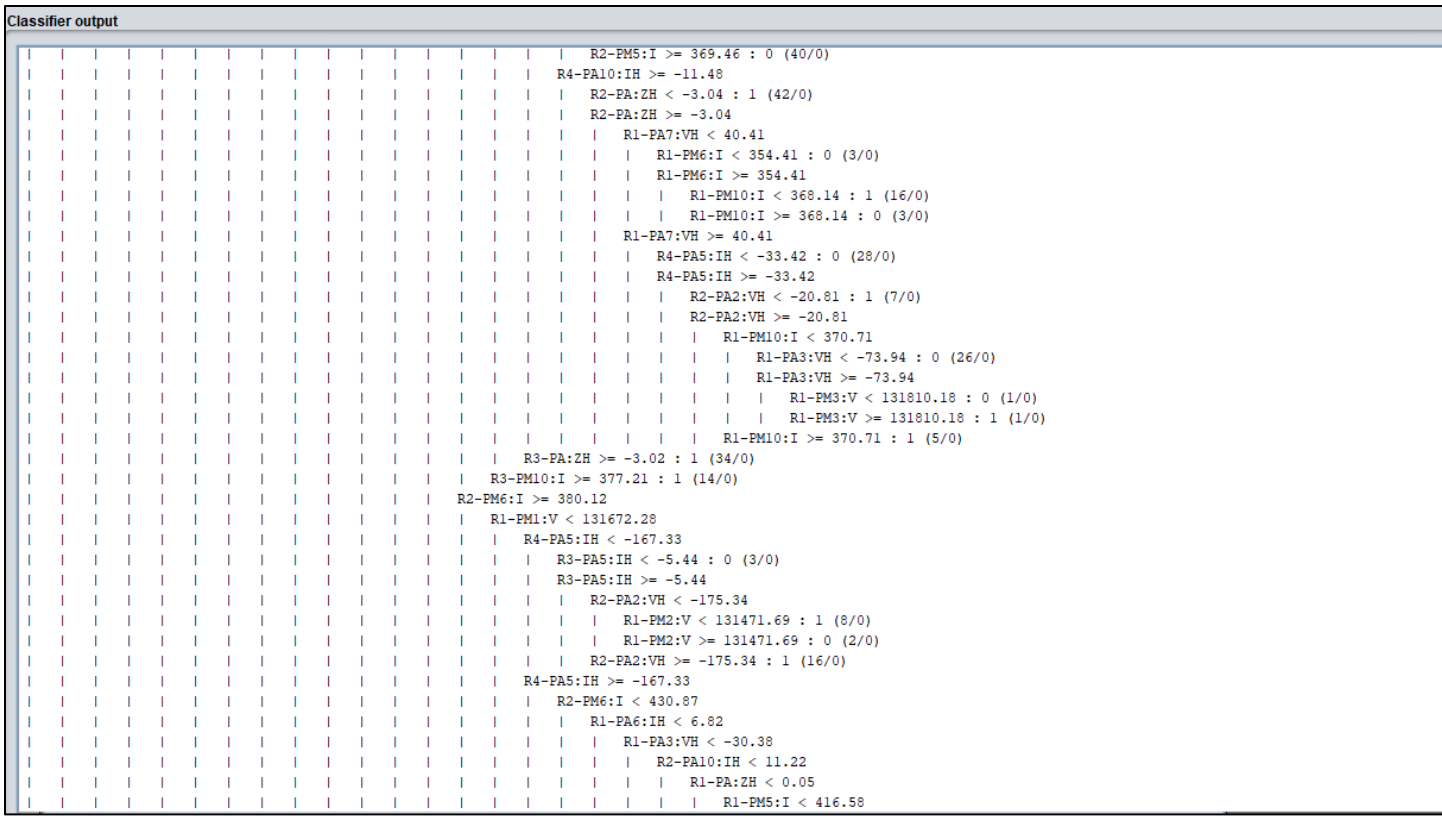

Classifier output
R2-PM3:V < 128749.61 : 1 (2/0)
R2-PM3:V >= 128749.61
R1-PA:ZH < 0.06 : 1 (2/0)
R1-PA:ZH >= 0.06
R3-PM4:I < 506.67
R1-PA5:IH < 84.21 : 0 (2/0)
R1-PA5:IH >= 84.21 : 1 (1/0)
R3-PM4:I >= 506.67 : 0 (7/0)
R3-PA10:IH >= 29.51 : 1 (26/0)
R1-PA3:VH >= 5.86
R1-PM4:I < 493.66
R2-PA10:IH < 85.17 : 0 (3/0)
R2-PA10:IH >= 85.17 : 1 (13/0)
R1-PM4:I >= 493.66 : 0 (26/0)
R1-PA4:IH >= -38.88
R1-PA7:VH < -29.28 : 1 (21/0)
R1-PA7:VH >= -29.28
R1-PM6:I < 507.31
R4-PM7:V < 130694.42
R1-PA2:VH < -30.34
R2-PM1:V < 128824.49 : 1 (6/0)
R2-PM1:V >= 128824.49 : 0 (8/0)
R1-PA2:VH >= -30.34 : 0 (20/0)
R4-PM7:V >= 130694.42
R1-PA2:VH < -43.07 : 0 (9/0)
R1-PA2:VH >= -43.07
R2-PA11:IH < -15.24 : 0 (3/0)
R2-PA11:IH >= -15.24
R1-PA10:IH < 107.34 : 1 (23/0)
R1-PA10:IH >= 107.34
R1-PA5:IH < 44.94 : 0 (8/0)
R1-PA5:IH >= 44.94 : 1 (14/0)
R1-PM6:I >= 507.31
R3-PA5:IH < 145.78 : 1 (41/0)
R3-PA5:IH >= 145.78
R1-PM1:V < 130719.49 : 0 (7/0)
R1-PM1:V >= 130719.49 : 1 (5/0)
R4-PA:Z >= 7.82

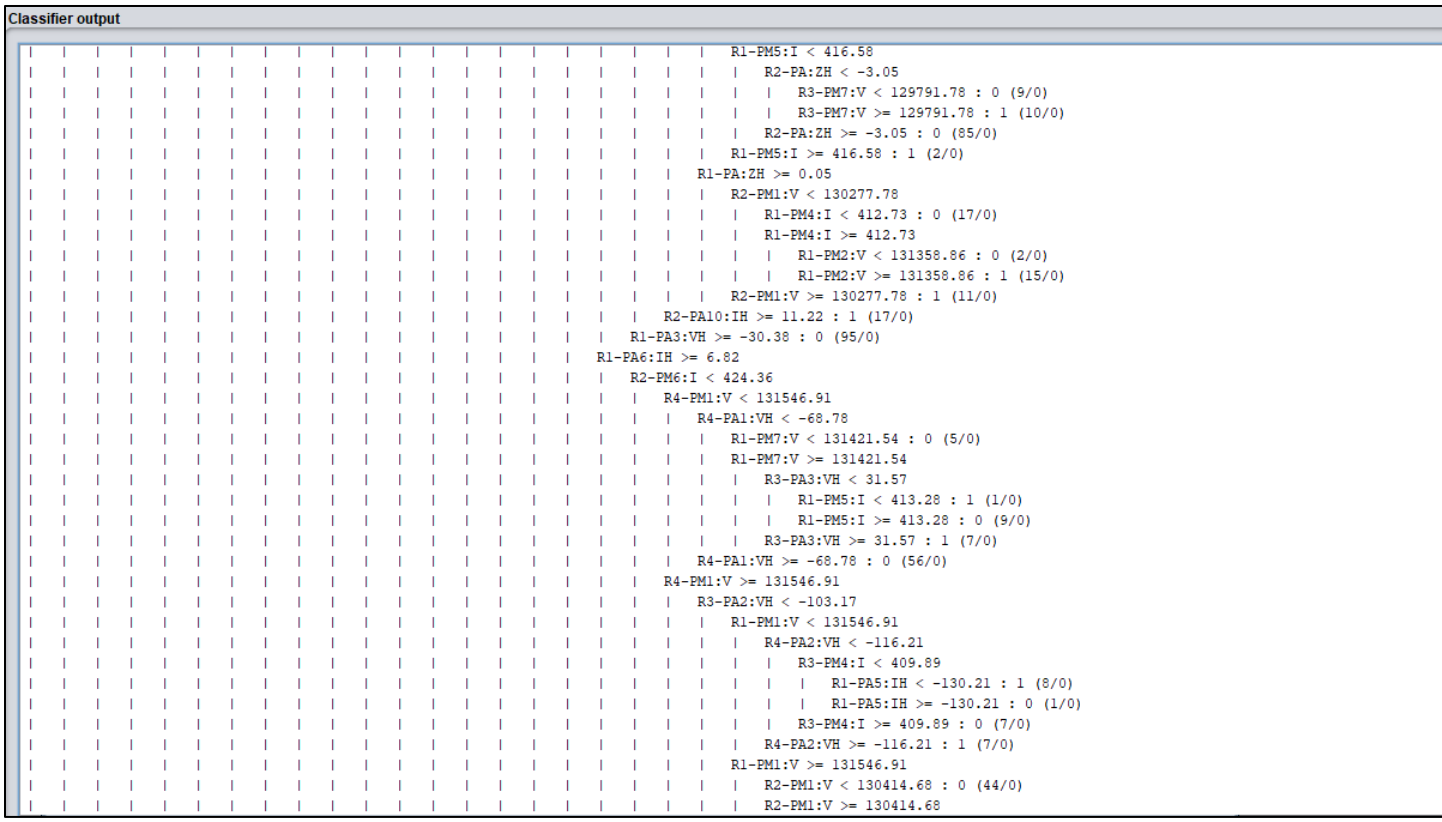

Classifier output
| | | | | | | | | | | | | | | | | | R4-PA:Z >= 7.82
| | | | | | | | | | | | | | | | | | | R2-PA:ZH < -3.07
| | | | | | | | | | | | | | | | | | | | R3-PA6:IH < 114.87 : 1 (7/0)
| | | | | | | | | | | | | | | | | | | | R3-PA6:IH >= 114.87
| | | | | | | | | | | | | | | | | | | | | R1-PM7:V < 130970.23 : 1 (1/0)
| | | | | | | | | | | | | | | | | | | | | R1-PM7:V >= 130970.23 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | R2-PA:ZH >= -3.07 : 0 (33/0)
| | | | | | | | | | | | | | | | | | R1-PA3:VH >= 104.08 : 0 (32/0)
| | | | | | | | | | | | | | | | | R2-PA3:VH >= 122.78 : 1 (24/0)
| | | | | | | | | | | | | | | | R2-PA3:VH >= 159.8 : 0 (32/0)
| | | | | | | | | | | | | | | R1-PM1:V >= 131020.37
| | | | | | | | | | | | | | | | R1-PM10:I < 497.6 : 1 (252/0)
| | | | | | | | | | | | | | | | R1-PM10:I >= 497.6 : 0 (2/0)
| | | | | | | | | | | | | | R2-PM7:V >= 129181.27
| | | | | | | | | | | | | | | R3-PA:ZH < -3.06
| | | | | | | | | | | | | | | | R4-PA10:IH < 46.58
| | | | | | | | | | | | | | | | | R2-PA12:IH < -26.41
| | | | | | | | | | | | | | | | | | R1-PA11:IH < -62.06
| | | | | | | | | | | | | | | | | | | R1-PA10:IH < -137.54 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | R1-PA10:IH >= -137.54 : 1 (1/0)
| | | | | | | | | | | | | | | | | | R1-PA11:IH >= -62.06 : 0 (8/0)
| | | | | | | | | | | | | | | | | R2-PA12:IH >= -26.41
| | | | | | | | | | | | | | | | | | R1-PM3:V < 131421.54
| | | | | | | | | | | | | | | | | | | R3-PA6:IH < -169.05 : 0 (7/0)
| | | | | | | | | | | | | | | | | | | R3-PA6:IH >= -169.05
| | | | | | | | | | | | | | | | | | | | R4-PM5:I < 431.86 : 1 (26/0)
| | | | | | | | | | | | | | | | | | | | R4-PM5:I >= 431.86
| | | | | | | | | | | | | | | | | | | | | R1-PA6:IH < 98.25
| | | | | | | | | | | | | | | | | | | | | | R3-PA6:IH < 136.77
| | | | | | | | | | | | | | | | | | | | | | | R1-PM10:I < 476.82
| | | | | | | | | | | | | | | | | | | | | | | | R3-PA3:VH < -58.99
| | | | | | | | | | | | | | | | | | | | | | | | | R1-PA5:IH < 61.87 : 1 (7/0)
| | | | | | | | | | | | | | | | | | | | | | | | | R1-PA5:IH >= 61.87 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | | | | | | R3-PA3:VH >= -58.99
| | | | | | | | | | | | | | | | | | | | | | | | | | R2-PM3:V < 129768.33 : 1 (52/0)
| | | | | | | | | | | | | | | | | | | | | | | | | | R2-PM3:V >= 129768.33 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | | | | | | | R1-PM10:I >= 476.82 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | | | | | | R3-PA6:IH >= 136.77 : 0 (6/0)

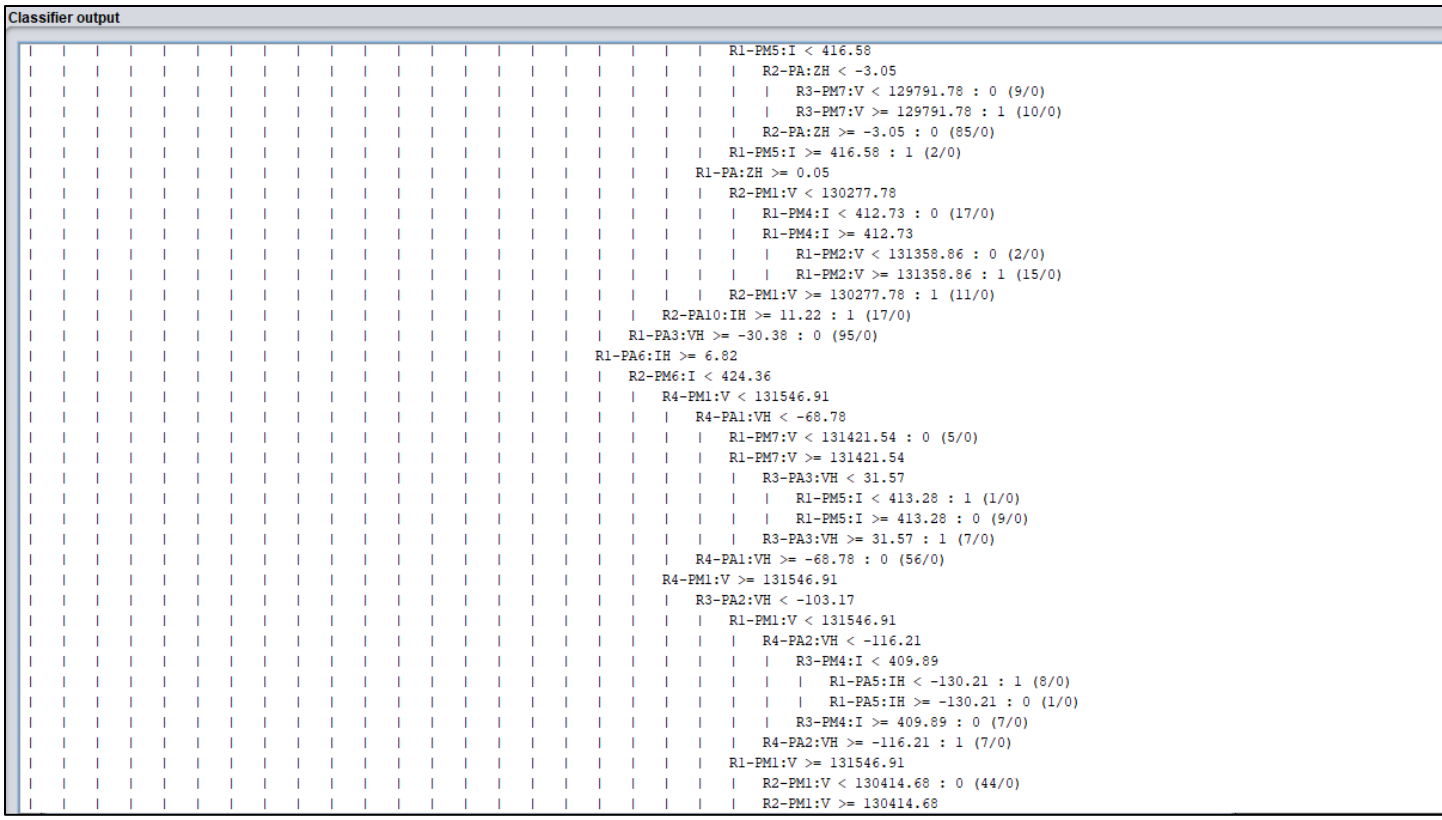

Classifier output
| | | | | | | | | | | | | | | | R3-PA6:IH >= 136.77 : 0 (6/0)
| | | | | | | | | | | | | | | R1-PA6:IH >= 98.25 : 0 (5/0)
| | | | | | | | | | | | | R1-PM3:V >= 131421.54
| | | | | | | | | | | | | | R1-PA1:VH < -98.45
| | | | | | | | | | | | | | | R4-PA3:VH < 19.26
| | | | | | | | | | | | | | | | R4-PA:ZH < 0.03 : 1 (33/0)
| | | | | | | | | | | | | | | | R4-PA:ZH >= 0.03
| | | | | | | | | | | | | | | | | R1-PA5:IH < 80.23 : 0 (2/0)
| | | | | | | | | | | | | | | | | R1-PA5:IH >= 80.23
| | | | | | | | | | | | | | | | | | R4-PM5:I < 406.69 : 0 (1/0)
| | | | | | | | | | | | | | | | | | R4-PM5:I >= 406.69 : 1 (26/0)
| | | | | | | | | | | | | | | R4-PA3:VH >= 19.26 : 0 (3/0)
| | | | | | | | | | | | | | R1-PA1:VH >= -98.45
| | | | | | | | | | | | | | | R1-PM11:I < 3.2 : 1 (120/0)
| | | | | | | | | | | | | | | R1-PM11:I >= 3.2
| | | | | | | | | | | | | | | | R2-PM1:V < 130018.51 : 1 (5/0)
| | | | | | | | | | | | | | | | R2-PM1:V >= 130018.51 : 0 (1/0)
| | | | | | | | | | | R4-PA10:IH >= 46.58
| | | | | | | | | | | | R3-PM1:V < 129039.58 : 0 (34/0)
| | | | | | | | | | | | R3-PM1:V >= 129039.58
| | | | | | | | | | | | | R2-PA1:VH < 66.73 : 0 (9/0)
| | | | | | | | | | | | | R2-PA1:VH >= 66.73
| | | | | | | | | | | | | | R4-PM7:V < 131220.96
| | | | | | | | | | | | | | | R2-PA2:VH < 21.87 : 0 (2/0)
| | | | | | | | | | | | | | | R2-PA2:VH >= 21.87
| | | | | | | | | | | | | | | | R3-PM2:V < 129290.32 : 1 (5/0)
| | | | | | | | | | | | | | | | R3-PM2:V >= 129290.32 : 0 (2/0)
| | | | | | | | | | | | | | R4-PM7:V >= 131220.96
| | | | | | | | | | | | | | | R2-PM4:I < 432.15 : 1 (40/0)
| | | | | | | | | | | | | | | R2-PM4:I >= 432.15
| | | | | | | | | | | | | | | | R1-PA:ZH < 0.06 : 0 (1/0)
| | | | | | | | | | | | | | | | R1-PA:ZH >= 0.06 : 1 (4/0)
| | | | | | | | | R3-PA:ZH >= -3.06
| | | | | | | | | | R1-PM1:V < 131321.25
| | | | | | | | | | | R3-PM6:I < 459.33
| | | | | | | | | | | | R2-PM2:V < 129448.26 : 0 (129/0)
| | | | | | | | | | | | R2-PM2:V >= 129448.26
| | | | | | | | | | | | | R4-PA1:VH < 120.02

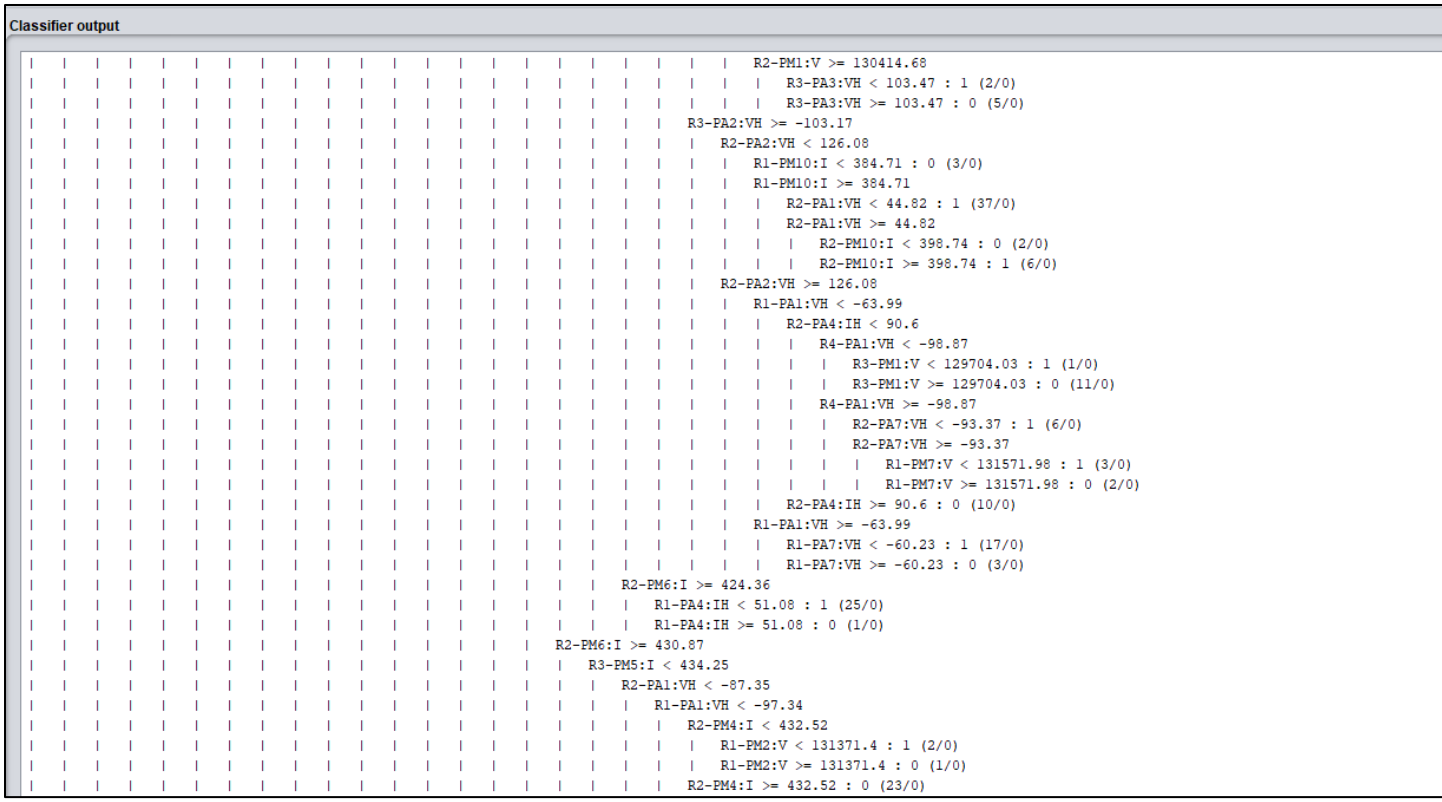

Classifier output
| | | | | | | | | | | | | | R4-PA1:VH < 120.02
| | | | | | | | | | | | | | | R2-PA2:VH < -170.32 : 1 (3/0)
| | | | | | | | | | | | | | | R2-PA2:VH >= -170.32
| | | | | | | | | | | | | | | | R1-PM6:I < 369.79 : 1 (1/0)
| | | | | | | | | | | | | | | | R1-PM6:I >= 369.79
| | | | | | | | | | | | | | | | | R2-PA1:VH < 175.53
| | | | | | | | | | | | | | | | | | R1-PA1:VH < 103.79 : 0 (115/0)
| | | | | | | | | | | | | | | | | | R1-PA1:VH >= 103.79
| | | | | | | | | | | | | | | | | | | R2-PM1:V < 129812.26 : 0 (9/0)
| | | | | | | | | | | | | | | | | | | R2-PM1:V >= 129812.26 : 1 (1/0)
| | | | | | | | | | | | | | | | | R2-PA1:VH >= 175.53 : 1 (1/0)
| | | | | | | | | | | | | | R4-PA1:VH >= 120.02
| | | | | | | | | | | | | | | R3-PM5:I < 435.98 : 0 (14/0)
| | | | | | | | | | | | | | | R3-PM5:I >= 435.98 : 1 (9/0)
| | | | | | | | | | | | | R3-PM6:I >= 459.33
| | | | | | | | | | | | | | R3-PA4:IH < 12.75
| | | | | | | | | | | | | | | R3-PA:ZH < -3.05
| | | | | | | | | | | | | | | | R2-PM4:I < 482.5
| | | | | | | | | | | | | | | | | R4-PA6:IH < -113.71
| | | | | | | | | | | | | | | | | | R2-PM2:V < 129074.07
| | | | | | | | | | | | | | | | | | | R1-PA1:VH < 121.48 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | R1-PA1:VH >= 121.48 : 1 (2/0)
| | | | | | | | | | | | | | | | | | R2-PM2:V >= 129074.07 : 0 (17/0)
| | | | | | | | | | | | | | | | | R4-PA6:IH >= -113.71 : 0 (66/0)
| | | | | | | | | | | | | | | | R2-PM4:I >= 482.5 : 1 (4/0)
| | | | | | | | | | | | | | | R3-PA:ZH >= -3.05 : 1 (7/0)
| | | | | | | | | | | | | | R3-PA4:IH >= 12.75
| | | | | | | | | | | | | | | R3-PA2:VH < -156.11 : 0 (8/0)
| | | | | | | | | | | | | | | R3-PA2:VH >= -156.11
| | | | | | | | | | | | | | | | R2-PA4:IH < 150.56
| | | | | | | | | | | | | | | | | R4-PM7:V < 130907.54 : 0 (5/0)
| | | | | | | | | | | | | | | | | R4-PM7:V >= 130907.54
| | | | | | | | | | | | | | | | | | R1-PM4:I < 463.82 : 1 (19/0)
| | | | | | | | | | | | | | | | | | R1-PM4:I >= 463.82
| | | | | | | | | | | | | | | | | | | R2-PA7:VH < -52.1 : 0 (7/0)
| | | | | | | | | | | | | | | | | | | R2-PA7:VH >= -52.1 : 1 (16/0)
| | | | | | | | | | | | | | | | R2-PA4:IH >= 150.56 : 0 (3/0)
| | | | | | | | | | | R1-PM1:V >= 131321.25

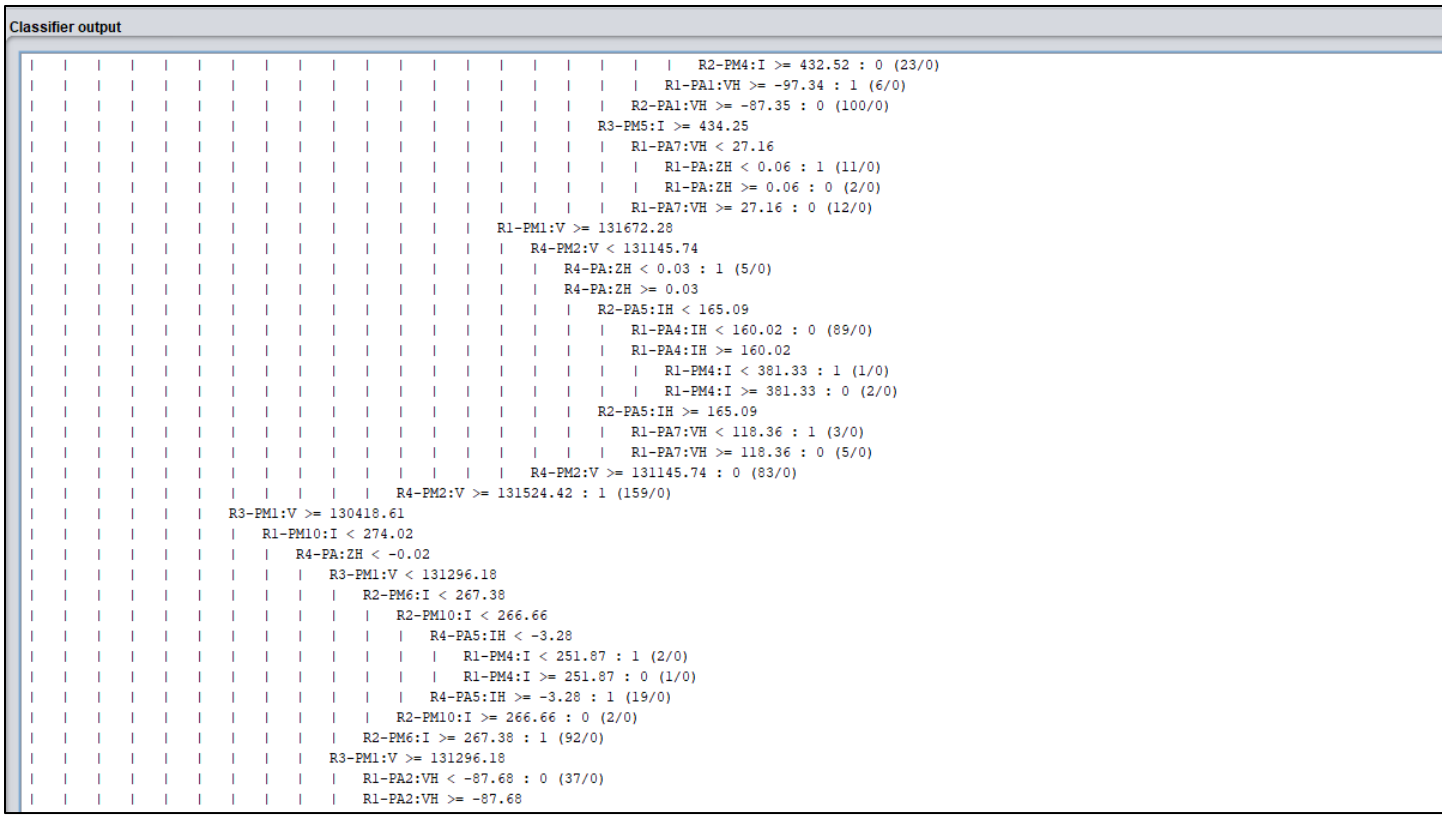

Classifier output
R1-PM1:V >= 131321.25
R4-PM2:V < 131524.42
R1-PA:ZH < 0.02
R3-PA6:IH < -113.68
R4-PM5:I < 370.98 : 0 (3/0)
R4-PM5:I >= 370.98 : 1 (2/0)
R3-PA6:IH >= -113.68
R1-PA3:VH < -26.76 : 1 (46/0)
R1-PA3:VH >= -26.76
R3:F < 60
R2-PA10:IH < 42.05 : 0 (1/0)
R2-PA10:IH >= 42.05 : 1 (19/0)
R3:F >= 60 : 0 (2/0)
R1-PA:ZH >= 0.02
R2-PM6:I < 380.12
R3-PM10:I < 377.21
R3-PA:ZH < -3.02
R4-PA10:IH < -11.48
R2-PM5:I < 369.46
R1-PM5:I < 348 : 0 (23/0)
R1-PM5:I >= 348
R2-PA5:IH < -62.21
R2-PA10:IH < 3.39
R1-PA2:VH < 72.47 : 0 (5/0)
R1-PA2:VH >= 72.47 : 1 (1/0)
R2-PA10:IH >= 3.39 : 0 (15/0)
R2-PA5:IH >= -62.21
R2-PA7:VH < -31.13
R3-PA6:IH < -109.52
R2-PA3:VH < 30.86 : 1 (4/0)
R2-PA3:VH >= 30.86
R2-PM10:I < 361.41 : 0 (12/0)
R2-PM10:I >= 361.41 : 1 (4/0)
R3-PA6:IH >= -109.52
R4-PM3:V < 131835.25 : 0 (1/0)
R4-PM3:V >= 131835.25 : 1 (12/0)
R2-PA7:VH >= -31.13 : 0 (10/0)
R2-PM5:I >= 369.46 : 0 (40/0)

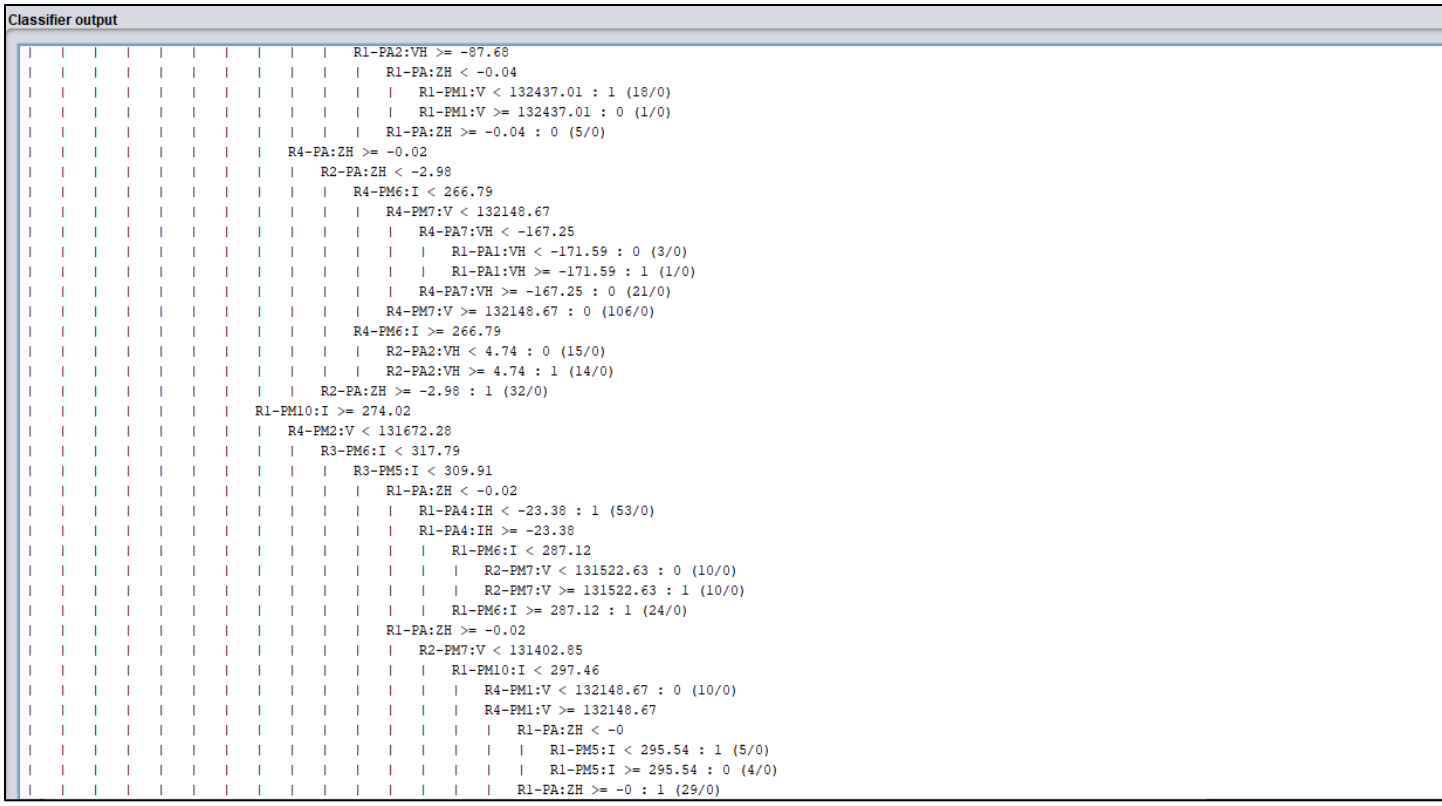

Classifier output
R2-PM5:I >= 369.46 : 0 (40/0)
R4-PA10:IH >= -11.48
R2-PA:ZH < -3.04 : 1 (42/0)
R2-PA:ZH >= -3.04
R1-PA7:VH < 40.41
R1-PM6:I < 354.41 : 0 (3/0)
R1-PM6:I >= 354.41
R1-PM10:I < 368.14 : 1 (16/0)
R1-PM10:I >= 368.14 : 0 (3/0)
R1-PA7:VH >= 40.41
R4-PA5:IH < -33.42 : 0 (28/0)
R4-PA5:IH >= -33.42
R2-PA2:VH < -20.81 : 1 (7/0)
R2-PA2:VH >= -20.81
R1-PM10:I < 370.71
R1-PA3:VH < -73.94 : 0 (26/0)
R1-PA3:VH >= -73.94
R1-PM3:V < 131810.18 : 0 (1/0)
R1-PM3:V >= 131810.18 : 1 (1/0)
R1-PM10:I >= 370.71 : 1 (5/0)
R3-PA:ZH >= -3.02 : 1 (34/0)
R3-PM10:I >= 377.21 : 1 (14/0)
R2-PM6:I >= 380.12
R1-PM1:V < 131672.28
R4-PA5:IH < -167.33
R3-PA5:IH < -5.44 : 0 (3/0)
R3-PA5:IH >= -5.44
R2-PA2:VH < -175.34
R1-PM2:V < 131471.69 : 1 (8/0)
R1-PM2:V >= 131471.69 : 0 (2/0)
R2-PA2:VH >= -175.34 : 1 (16/0)
R4-PA5:IH >= -167.33
R2-PM6:I < 430.87
R1-PA6:IH < 6.82
R1-PA3:VH < -30.38
R2-PA10:IH < 11.22
R1-PA:ZH < 0.05
R1-PM5:I < 416.58

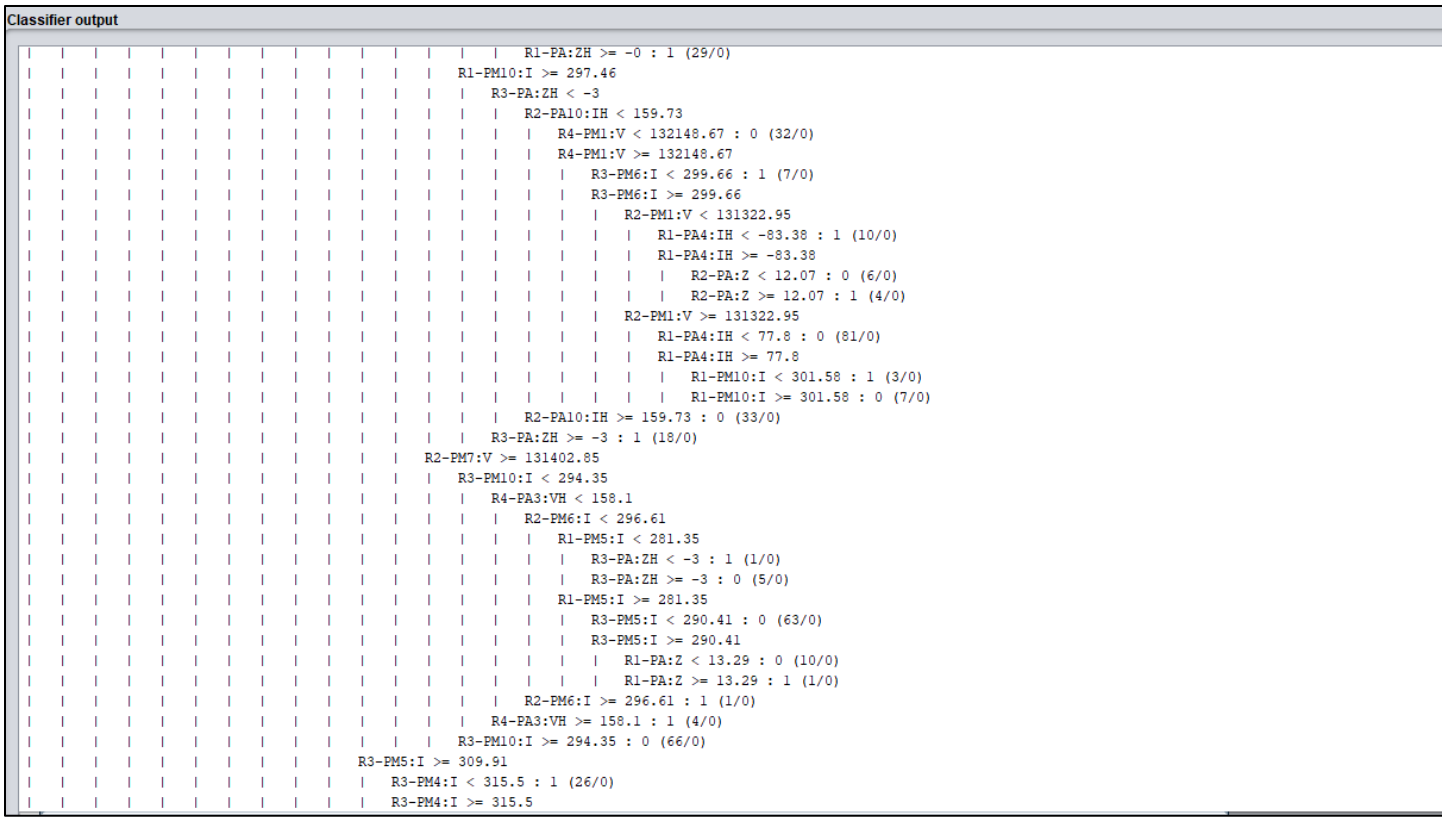

Classifier output
R1-PM5:I < 416.58
R2-PA:ZH < -3.05
R3-PM7:V < 129791.78 : 0 (9/0)
R3-PM7:V >= 129791.78 : 1 (10/0)
R2-PA:ZH >= -3.05 : 0 (85/0)
R1-PM5:I >= 416.58 : 1 (2/0)
R1-PA:ZH >= 0.05
R2-PM1:V < 130277.78
R1-PM4:I < 412.73 : 0 (17/0)
R1-PM4:I >= 412.73
R1-PM2:V < 131358.86 : 0 (2/0)
R1-PM2:V >= 131358.86 : 1 (15/0)
R2-PM1:V >= 130277.78 : 1 (11/0)
R2-PA10:IH >= 11.22 : 1 (17/0)
R1-PA3:VH >= -30.38 : 0 (95/0)
R1-PA6:IH >= 6.82
R2-PM6:I < 424.36
R4-PM1:V < 131546.91
R4-PA1:VH < -68.78
R1-PM7:V < 131421.54 : 0 (5/0)
R1-PM7:V >= 131421.54
R3-PA3:VH < 31.57
R1-PM5:I < 413.28 : 1 (1/0)
R1-PM5:I >= 413.28 : 0 (9/0)
R3-PA3:VH >= 31.57 : 1 (7/0)
R4-PA1:VH >= -68.78 : 0 (56/0)
R4-PM1:V >= 131546.91
R3-PA2:VH < -103.17
R1-PM1:V < 131546.91
R4-PA2:VH < -116.21
R3-PM4:I < 409.89
R1-PA5:IH < -130.21 : 1 (8/0)
R1-PA5:IH >= -130.21 : 0 (1/0)
R3-PM4:I >= 409.89 : 0 (7/0)
R4-PA2:VH >= -116.21 : 1 (7/0)
R1-PM1:V >= 131546.91
R2-PM1:V < 130414.68 : 0 (44/0)
R2-PM1:V >= 130414.68

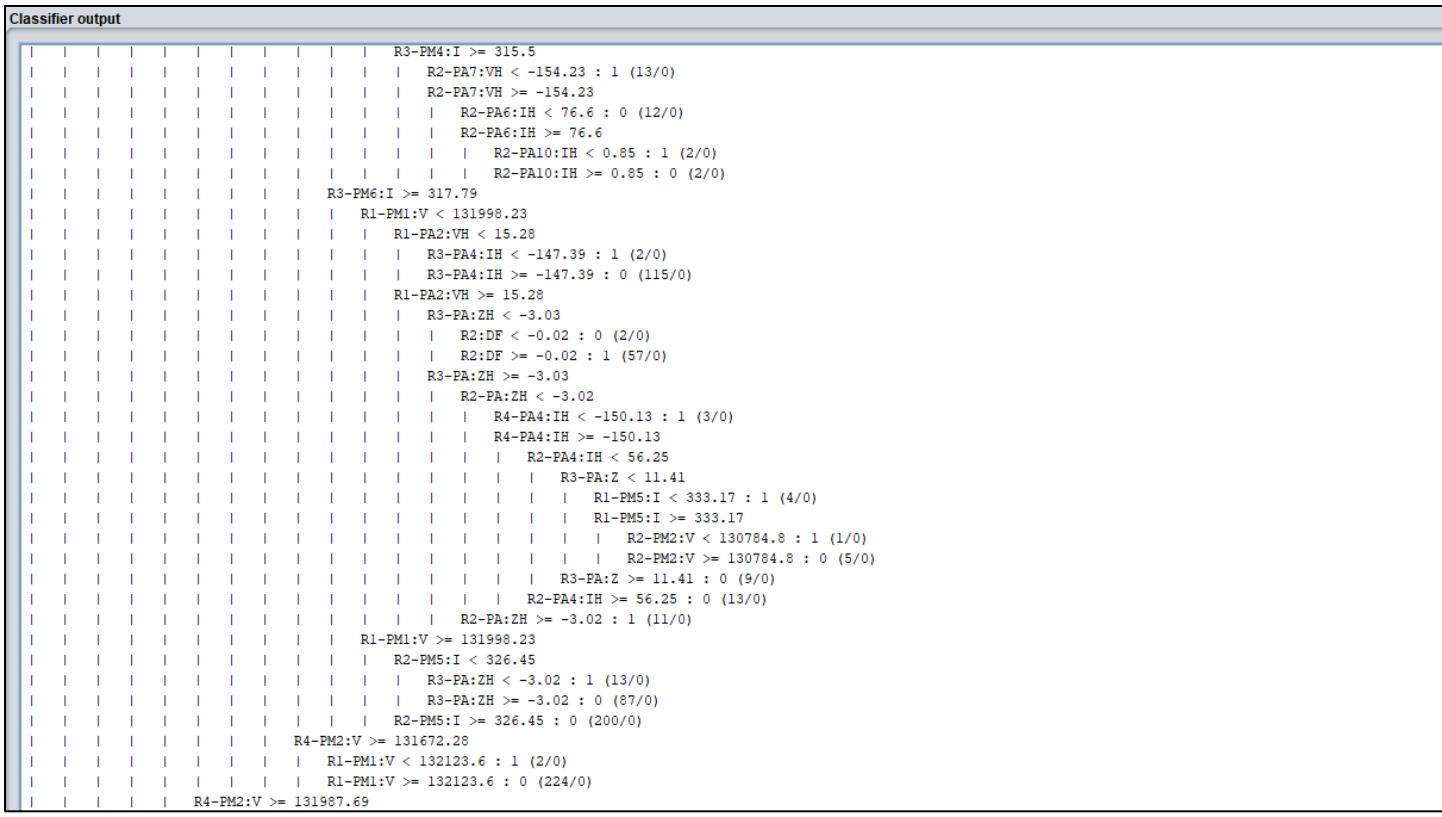

Classifier output
R1-PM5:I < 416.58
R2-PA:ZH < -3.05
R3-PM7:V < 129791.78 : 0 (9/0)
R3-PM7:V >= 129791.78 : 1 (10/0)
R2-PA:ZH >= -3.05 : 0 (85/0)
R1-PM5:I >= 416.58 : 1 (2/0)
R1-PA:ZH >= 0.05
R2-PM1:V < 130277.78
R1-PM4:I < 412.73 : 0 (17/0)
R1-PM4:I >= 412.73
R1-PM2:V < 131358.86 : 0 (2/0)
R1-PM2:V >= 131358.86 : 1 (15/0)
R2-PM1:V >= 130277.78 : 1 (11/0)
R2-PA10:IH >= 11.22 : 1 (17/0)
R1-PA3:VH >= -30.38 : 0 (95/0)
R1-PA6:IH >= 6.82
R2-PM6:I < 424.36
R4-PM1:V < 131546.91
R4-PA1:VH < -68.78
R1-PM7:V < 131421.54 : 0 (5/0)
R1-PM7:V >= 131421.54
R3-PA3:VH < 31.57
R1-PM5:I < 413.28 : 1 (1/0)
R1-PM5:I >= 413.28 : 0 (9/0)
R3-PA3:VH >= 31.57 : 1 (7/0)
R4-PA1:VH >= -68.78 : 0 (56/0)
R4-PM1:V >= 131546.91
R3-PA2:VH < -103.17
R1-PM1:V < 131546.91
R4-PA2:VH < -116.21
R3-PM4:I < 409.89
R1-PA5:IH < -130.21 : 1 (8/0)
R1-PA5:IH >= -130.21 : 0 (1/0)
R3-PM4:I >= 409.89 : 0 (7/0)
R4-PA2:VH >= -116.21 : 1 (7/0)
R1-PM1:V >= 131546.91
R2-PM1:V < 130414.68 : 0 (44/0)
R2-PM1:V >= 130414.68

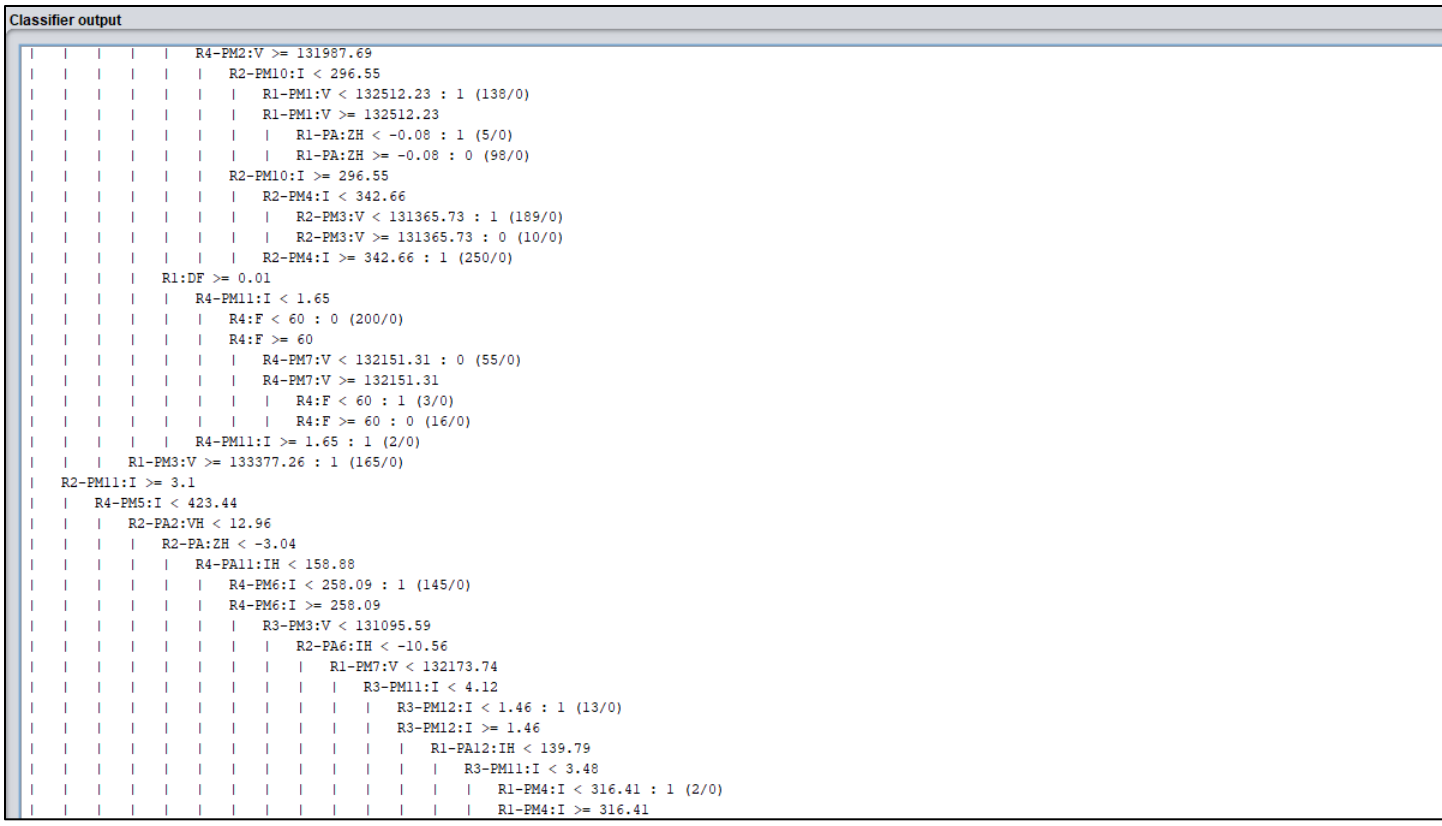

Classifier output
| | | | | | | | | | | | | | | | | | | | | | R2-PM1:V >= 130414.68
| | | | | | | | | | | | | | | | | | | | | | | R3-PA3:VH < 103.47 : 1 (2/0)
| | | | | | | | | | | | | | | | | | | | | | | R3-PA3:VH >= 103.47 : 0 (5/0)
| | | | | | | | | | | | | | | | | | | | | R3-PA2:VH >= -103.17
| | | | | | | | | | | | | | | | | | | | | | R2-PA2:VH < 126.08
| | | | | | | | | | | | | | | | | | | | | | | R1-PM10:I < 384.71 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | | | | | R1-PM10:I >= 384.71
| | | | | | | | | | | | | | | | | | | | | | | | R2-PA1:VH < 44.82 : 1 (37/0)
| | | | | | | | | | | | | | | | | | | | | | | | R2-PA1:VH >= 44.82
| | | | | | | | | | | | | | | | | | | | | | | | | R2-PM10:I < 398.74 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | | | | | | | R2-PM10:I >= 398.74 : 1 (6/0)
| | | | | | | | | | | | | | | | | | | | | | R2-PA2:VH >= 126.08
| | | | | | | | | | | | | | | | | | | | | | | R1-PA1:VH < -63.99
| | | | | | | | | | | | | | | | | | | | | | | | R2-PA4:IH < 90.6
| | | | | | | | | | | | | | | | | | | | | | | | | R4-PA1:VH < -98.87
| | | | | | | | | | | | | | | | | | | | | | | | | | R3-PM1:V < 129704.03 : 1 (1/0)
| | | | | | | | | | | | | | | | | | | | | | | | | | R3-PM1:V >= 129704.03 : 0 (11/0)
| | | | | | | | | | | | | | | | | | | | | | | | | R4-PA1:VH >= -98.87
| | | | | | | | | | | | | | | | | | | | | | | | | | R2-PA7:VH < -93.37 : 1 (6/0)
| | | | | | | | | | | | | | | | | | | | | | | | | | R2-PA7:VH >= -93.37
| | | | | | | | | | | | | | | | | | | | | | | | | | | R1-PM7:V < 131571.98 : 1 (3/0)
| | | | | | | | | | | | | | | | | | | | | | | | | | | R1-PM7:V >= 131571.98 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | | | | | | R2-PA4:IH >= 90.6 : 0 (10/0)
| | | | | | | | | | | | | | | | | | | | | | | R1-PA1:VH >= -63.99
| | | | | | | | | | | | | | | | | | | | | | | | R1-PA7:VH < -60.23 : 1 (17/0)
| | | | | | | | | | | | | | | | | | | | | | | | R1-PA7:VH >= -60.23 : 0 (3/0)
| | | | | | | | | | | | | | | | | | | | | | R2-PM6:I >= 424.36
| | | | | | | | | | | | | | | | | | | | | | | R1-PA4:IH < 51.08 : 1 (25/0)
| | | | | | | | | | | | | | | | | | | | | | | R1-PA4:IH >= 51.08 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | | | R2-PM6:I >= 430.87
| | | | | | | | | | | | | | | | | | | | | | R3-PM5:I < 434.25
| | | | | | | | | | | | | | | | | | | | | | | R2-PA1:VH < -87.35
| | | | | | | | | | | | | | | | | | | | | | | | R1-PA1:VH < -97.34
| | | | | | | | | | | | | | | | | | | | | | | | | R2-PM4:I < 432.52
| | | | | | | | | | | | | | | | | | | | | | | | | | R1-PM2:V < 131371.4 : 1 (2/0)
| | | | | | | | | | | | | | | | | | | | | | | | | | R1-PM2:V >= 131371.4 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | | | | | | | R2-PM4:I >= 432.52 : 0 (23/0)

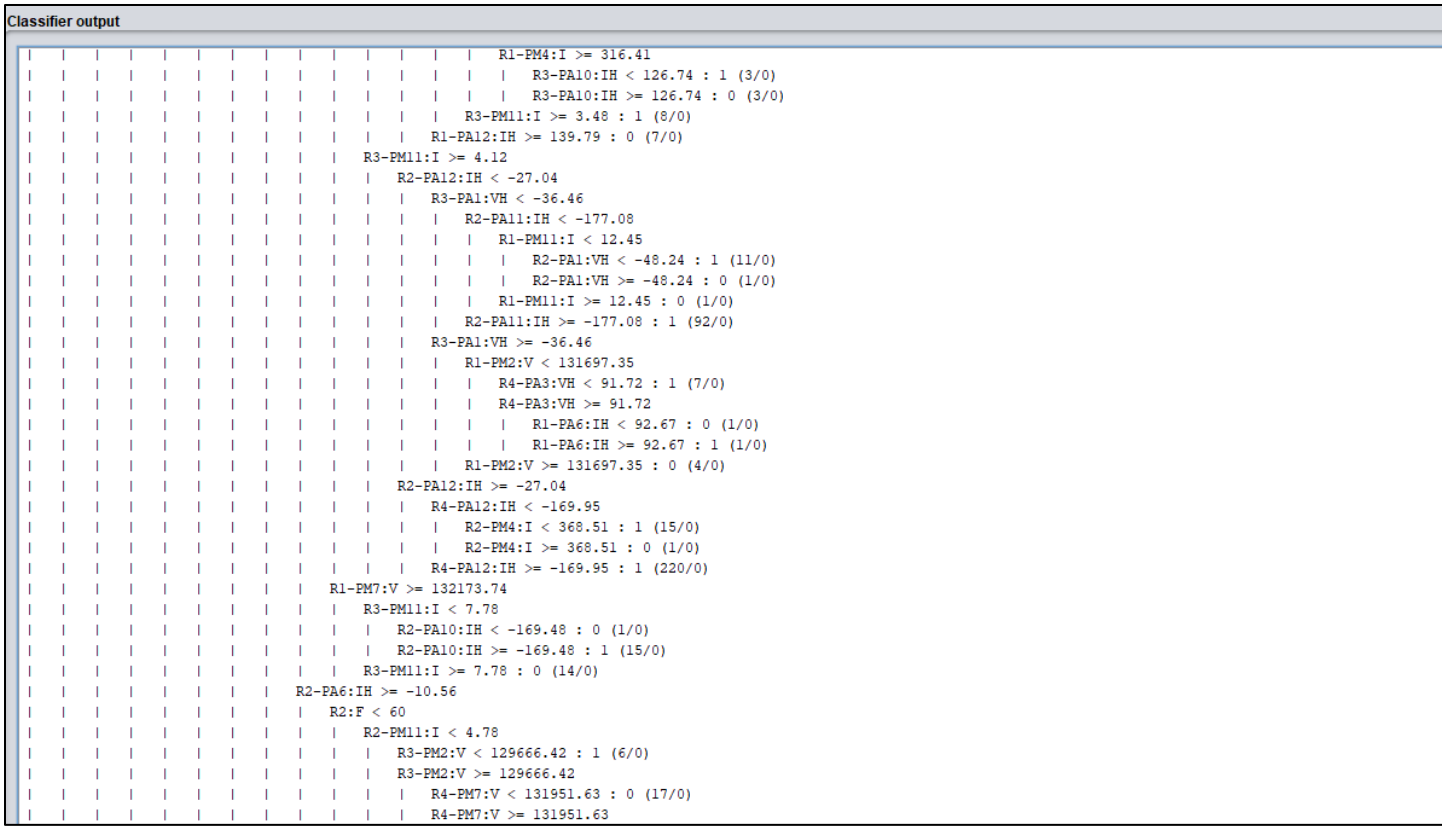

Classifier output
| | | | | | | | | | | | | | | | | | | R2-PM4:I >= 432.52 : 0 (23/0)
| | | | | | | | | | | | | | | | | | | R1-PA1:VH >= -97.34 : 1 (6/0)
| | | | | | | | | | | | | | | | | | | R2-PA1:VH >= -87.35 : 0 (100/0)
| | | | | | | | | | | | | | | | | | R3-PM5:I >= 434.25
| | | | | | | | | | | | | | | | | | | R1-PA7:VH < 27.16
| | | | | | | | | | | | | | | | | | | | R1-PA:ZH < 0.06 : 1 (11/0)
| | | | | | | | | | | | | | | | | | | | R1-PA:ZH >= 0.06 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | R1-PA7:VH >= 27.16 : 0 (12/0)
| | | | | | | | | | | | | | | | | | R1-PM1:V >= 131672.28
| | | | | | | | | | | | | | | | | | R4-PM2:V < 131145.74
| | | | | | | | | | | | | | | | | | R4-PA:ZH < 0.03 : 1 (5/0)
| | | | | | | | | | | | | | | | | | R4-PA:ZH >= 0.03
| | | | | | | | | | | | | | | | | | | R2-PA5:IH < 165.09
| | | | | | | | | | | | | | | | | | | R1-PA4:IH < 160.02 : 0 (89/0)
| | | | | | | | | | | | | | | | | | | R1-PA4:IH >= 160.02
| | | | | | | | | | | | | | | | | | | | R1-PM4:I < 381.33 : 1 (1/0)
| | | | | | | | | | | | | | | | | | | | R1-PM4:I >= 381.33 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | R2-PA5:IH >= 165.09
| | | | | | | | | | | | | | | | | | | R1-PA7:VH < 118.36 : 1 (3/0)
| | | | | | | | | | | | | | | | | | | R1-PA7:VH >= 118.36 : 0 (5/0)
| | | | | | | | | | | | | | | | | | R4-PM2:V >= 131145.74 : 0 (83/0)
| | | | | | | | | | | | | | | | | R4-PM2:V >= 131524.42 : 1 (159/0)
| | | | | | | R3-PM1:V >= 130418.61
| | | | | | | | R1-PM10:I < 274.02
| | | | | | | | | R4-PA:ZH < -0.02
| | | | | | | | | | R3-PM1:V < 131296.18
| | | | | | | | | | | R2-PM6:I < 267.38
| | | | | | | | | | | | R2-PM10:I < 266.66
| | | | | | | | | | | | | | R4-PA5:IH < -3.28
| | | | | | | | | | | | | | | R1-PM4:I < 251.87 : 1 (2/0)
| | | | | | | | | | | | | | | R1-PM4:I >= 251.87 : 0 (1/0)
| | | | | | | | | | | | | | R4-PA5:IH >= -3.28 : 1 (19/0)
| | | | | | | | | | | | | R2-PM10:I >= 266.66 : 0 (2/0)
| | | | | | | | | | | | R2-PM6:I >= 267.38 : 1 (92/0)
| | | | | | | | | | | R3-PM1:V >= 131296.18
| | | | | | | | | | | | R1-PA2:VH < -87.68 : 0 (37/0)
| | | | | | | | | | | | R1-PA2:VH >= -87.68

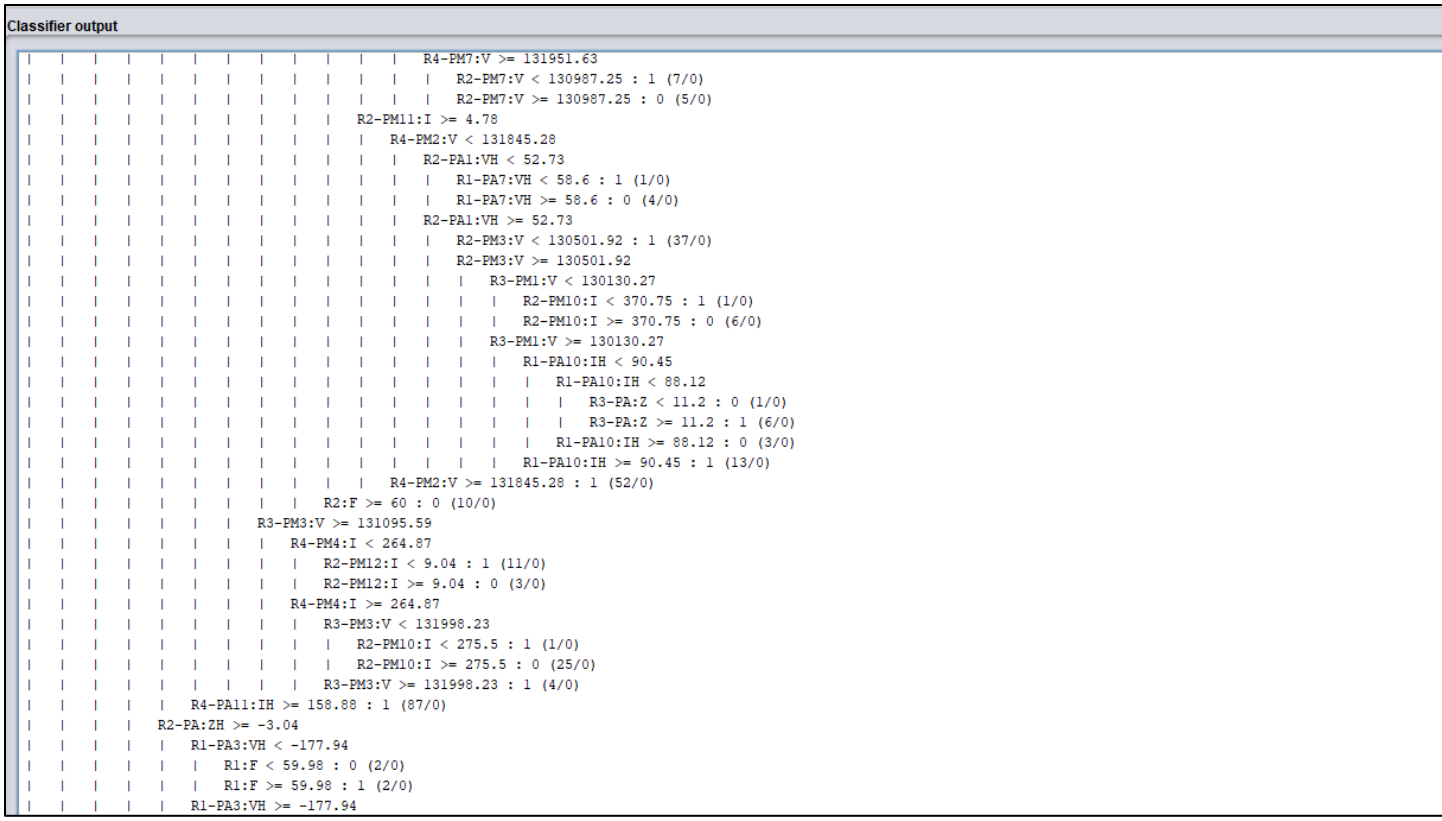

Classifier output
| | | | | | | | | | R1-PA2:VH >= -87.68
| | | | | | | | | | | R1-PA:ZH < -0.04
| | | | | | | | | | | | R1-PM1:V < 132437.01 : 1 (18/0)
| | | | | | | | | | | | R1-PM1:V >= 132437.01 : 0 (1/0)
| | | | | | | | | | | R1-PA:ZH >= -0.04 : 0 (5/0)
| | | | | | | | | R4-PA:ZH >= -0.02
| | | | | | | | | | R2-PA:ZH < -2.98
| | | | | | | | | | | R4-PM6:I < 266.79
| | | | | | | | | | | | R4-PM7:V < 132148.67
| | | | | | | | | | | | | R4-PA7:VH < -167.25
| | | | | | | | | | | | | | R1-PA1:VH < -171.59 : 0 (3/0)
| | | | | | | | | | | | | | R1-PA1:VH >= -171.59 : 1 (1/0)
| | | | | | | | | | | | | R4-PA7:VH >= -167.25 : 0 (21/0)
| | | | | | | | | | | | R4-PM7:V >= 132148.67 : 0 (106/0)
| | | | | | | | | | | R4-PM6:I >= 266.79
| | | | | | | | | | | | R2-PA2:VH < 4.74 : 0 (15/0)
| | | | | | | | | | | | R2-PA2:VH >= 4.74 : 1 (14/0)
| | | | | | | | | | R2-PA:ZH >= -2.98 : 1 (32/0)
| | | | | | | | R1-PM10:I >= 274.02
| | | | | | | | | R4-PM2:V < 131672.28
| | | | | | | | | | R3-PM6:I < 317.79
| | | | | | | | | | | R3-PM5:I < 309.91
| | | | | | | | | | | | R1-PA:ZH < -0.02
| | | | | | | | | | | | | R1-PA4:IH < -23.38 : 1 (53/0)
| | | | | | | | | | | | | R1-PA4:IH >= -23.38
| | | | | | | | | | | | | | R1-PM6:I < 287.12
| | | | | | | | | | | | | | | R2-PM7:V < 131522.63 : 0 (10/0)
| | | | | | | | | | | | | | | R2-PM7:V >= 131522.63 : 1 (10/0)
| | | | | | | | | | | | | | R1-PM6:I >= 287.12 : 1 (24/0)
| | | | | | | | | | | | | R1-PA:ZH >= -0.02
| | | | | | | | | | | | | | R2-PM7:V < 131402.85
| | | | | | | | | | | | | | | R1-PM10:I < 297.46
| | | | | | | | | | | | | | | | R4-PM1:V < 132148.67 : 0 (10/0)
| | | | | | | | | | | | | | | | R4-PM1:V >= 132148.67
| | | | | | | | | | | | | | | | | R1-PA:ZH < -0
| | | | | | | | | | | | | | | | | | R1-PM5:I < 295.54 : 1 (5/0)
| | | | | | | | | | | | | | | | | | R1-PM5:I >= 295.54 : 0 (4/0)
| | | | | | | | | | | | | | | | | R1-PA:ZH >= -0 : 1 (29/0)

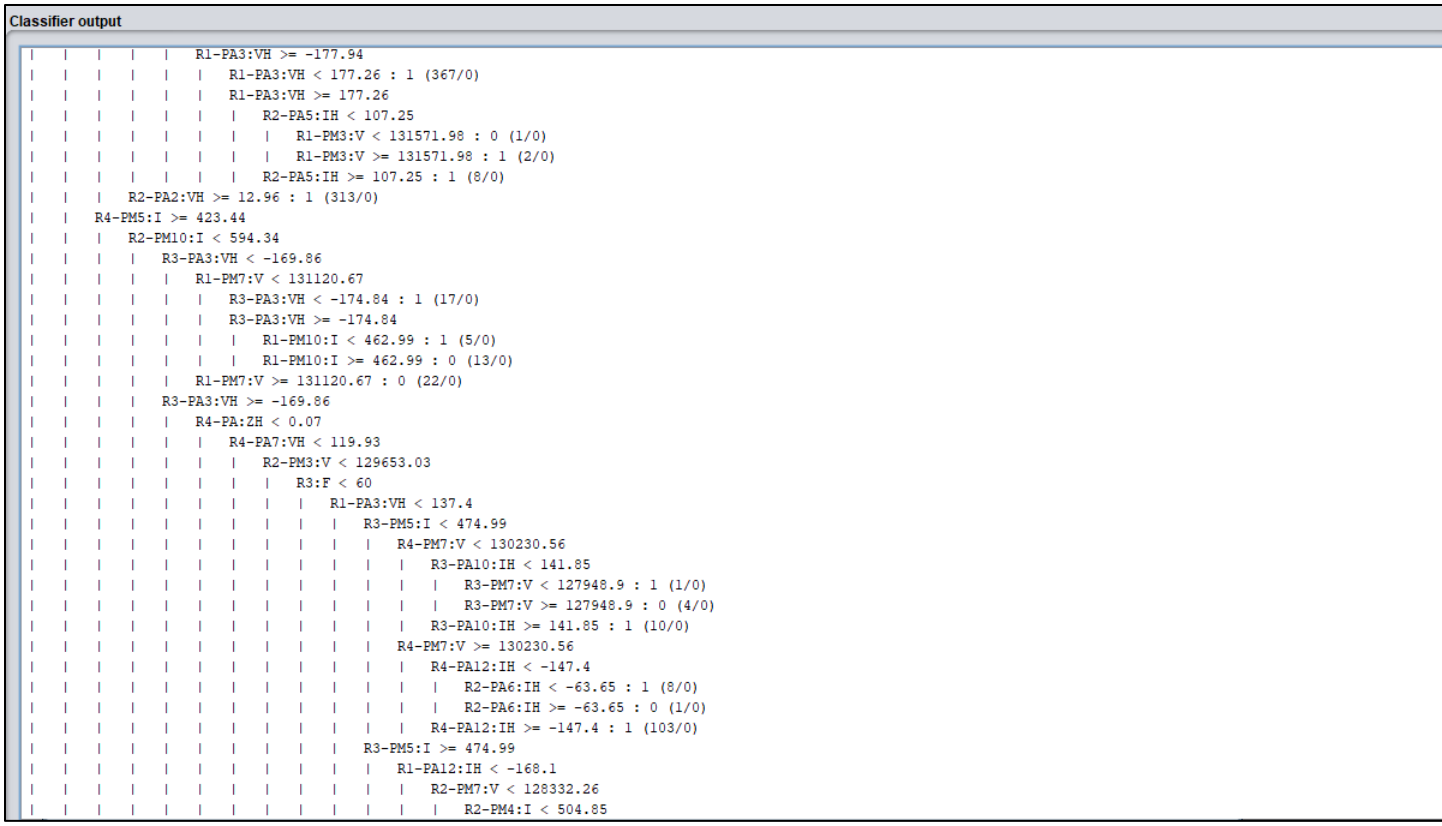

Classifier output
| | | | | | | | | | | | | | | R1-PA:ZH >= -0 : 1 (29/0)
| | | | | | | | | | | | | | R1-PM10:I >= 297.46
| | | | | | | | | | | | | | | R3-PA:ZH < -3
| | | | | | | | | | | | | | | | R2-PA10:IH < 159.73
| | | | | | | | | | | | | | | | | R4-PM1:V < 132148.67 : 0 (32/0)
| | | | | | | | | | | | | | | | | R4-PM1:V >= 132148.67
| | | | | | | | | | | | | | | | | | R3-PM6:I < 299.66 : 1 (7/0)
| | | | | | | | | | | | | | | | | | R3-PM6:I >= 299.66
| | | | | | | | | | | | | | | | | | | R2-PM1:V < 131322.95
| | | | | | | | | | | | | | | | | | | | R1-PA4:IH < -83.38 : 1 (10/0)
| | | | | | | | | | | | | | | | | | | | R1-PA4:IH >= -83.38
| | | | | | | | | | | | | | | | | | | | | R2-PA:Z < 12.07 : 0 (6/0)
| | | | | | | | | | | | | | | | | | | | | R2-PA:Z >= 12.07 : 1 (4/0)
| | | | | | | | | | | | | | | | | | | R2-PM1:V >= 131322.95
| | | | | | | | | | | | | | | | | | | | R1-PA4:IH < 77.8 : 0 (81/0)
| | | | | | | | | | | | | | | | | | | | R1-PA4:IH >= 77.8
| | | | | | | | | | | | | | | | | | | | | R1-PM10:I < 301.58 : 1 (3/0)
| | | | | | | | | | | | | | | | | | | | | R1-PM10:I >= 301.58 : 0 (7/0)
| | | | | | | | | | | | | | | | R2-PA10:IH >= 159.73 : 0 (33/0)
| | | | | | | | | | | | | | | R3-PA:ZH >= -3 : 1 (18/0)
| | | | | | | | | | | | | | R2-PM7:V >= 131402.85
| | | | | | | | | | | | | | | R3-PM10:I < 294.35
| | | | | | | | | | | | | | | | R4-PA3:VH < 158.1
| | | | | | | | | | | | | | | | | R2-PM6:I < 296.61
| | | | | | | | | | | | | | | | | | R1-PM5:I < 281.35
| | | | | | | | | | | | | | | | | | | R3-PA:ZH < -3 : 1 (1/0)
| | | | | | | | | | | | | | | | | | | R3-PA:ZH >= -3 : 0 (5/0)
| | | | | | | | | | | | | | | | | | R1-PM5:I >= 281.35
| | | | | | | | | | | | | | | | | | | R3-PM5:I < 290.41 : 0 (63/0)
| | | | | | | | | | | | | | | | | | | R3-PM5:I >= 290.41
| | | | | | | | | | | | | | | | | | | | R1-PA:Z < 13.29 : 0 (10/0)
| | | | | | | | | | | | | | | | | | | | R1-PA:Z >= 13.29 : 1 (1/0)
| | | | | | | | | | | | | | | | | | R2-PM6:I >= 296.61 : 1 (1/0)
| | | | | | | | | | | | | | | | R4-PA3:VH >= 158.1 : 1 (4/0)
| | | | | | | | | | | | | | | R3-PM10:I >= 294.35 : 0 (66/0)
| | | | | | | | | | | | | R3-PM5:I >= 309.91
| | | | | | | | | | | | | | R3-PM4:I < 315.5 : 1 (26/0)
| | | | | | | | | | | | | | R3-PM4:I >= 315.5

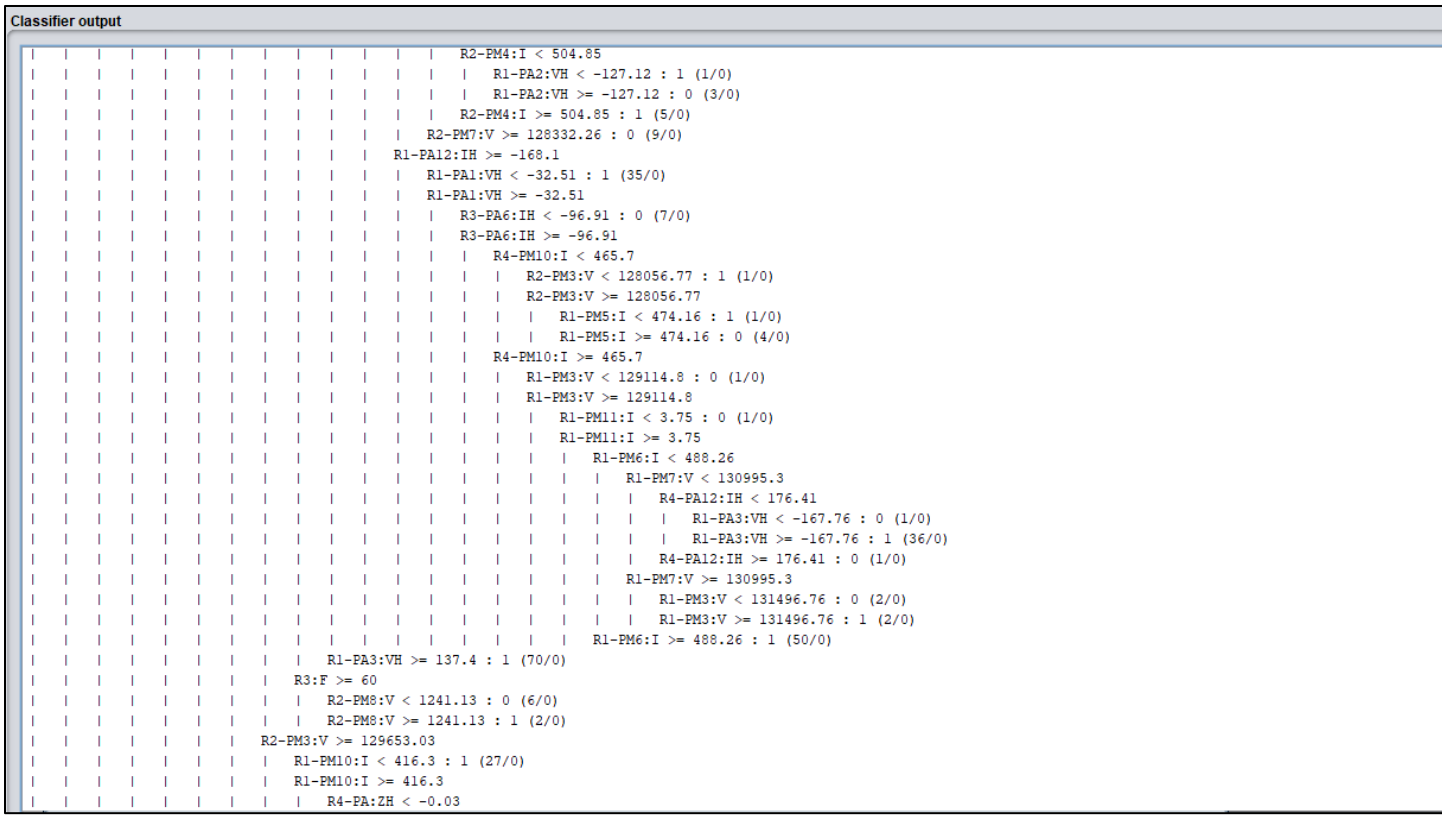

Classifier output
| | | | | | | | | | | R3-PM4:I >= 315.5
| | | | | | | | | | | | R2-PA7:VH < -154.23 : 1 (13/0)
| | | | | | | | | | | | R2-PA7:VH >= -154.23
| | | | | | | | | | | | | R2-PA6:IH < 76.6 : 0 (12/0)
| | | | | | | | | | | | | R2-PA6:IH >= 76.6
| | | | | | | | | | | | | | R2-PA10:IH < 0.85 : 1 (2/0)
| | | | | | | | | | | | | | R2-PA10:IH >= 0.85 : 0 (2/0)
| | | | | | | | | | R3-PM6:I >= 317.79
| | | | | | | | | | | R1-PM1:V < 131998.23
| | | | | | | | | | | R1-PA2:VH < 15.28
| | | | | | | | | | | | R3-PA4:IH < -147.39 : 1 (2/0)
| | | | | | | | | | | | R3-PA4:IH >= -147.39 : 0 (115/0)
| | | | | | | | | | | R1-PA2:VH >= 15.28
| | | | | | | | | | | | R3-PA:ZH < -3.03
| | | | | | | | | | | | | R2:DF < -0.02 : 0 (2/0)
| | | | | | | | | | | | | R2:DF >= -0.02 : 1 (57/0)
| | | | | | | | | | | | R3-PA:ZH >= -3.03
| | | | | | | | | | | | | R2-PA:ZH < -3.02
| | | | | | | | | | | | | | R4-PA4:IH < -150.13 : 1 (3/0)
| | | | | | | | | | | | | | R4-PA4:IH >= -150.13
| | | | | | | | | | | | | | | R2-PA4:IH < 56.25
| | | | | | | | | | | | | | | | R3-PA:Z < 11.41
| | | | | | | | | | | | | | | | | R1-PM5:I < 333.17 : 1 (4/0)
| | | | | | | | | | | | | | | | | R1-PM5:I >= 333.17
| | | | | | | | | | | | | | | | | | R2-PM2:V < 130784.8 : 1 (1/0)
| | | | | | | | | | | | | | | | | | R2-PM2:V >= 130784.8 : 0 (5/0)
| | | | | | | | | | | | | | | | R3-PA:Z >= 11.41 : 0 (9/0)
| | | | | | | | | | | | | | | R2-PA4:IH >= 56.25 : 0 (13/0)
| | | | | | | | | | | | | | R2-PA:ZH >= -3.02 : 1 (11/0)
| | | | | | | | | | | R1-PM1:V >= 131998.23
| | | | | | | | | | | | R2-PM5:I < 326.45
| | | | | | | | | | | | | R3-PA:ZH < -3.02 : 1 (13/0)
| | | | | | | | | | | | | R3-PA:ZH >= -3.02 : 0 (87/0)
| | | | | | | | | | | | R2-PM5:I >= 326.45 : 0 (200/0)
| | | | | | | | | R4-PM2:V >= 131672.28
| | | | | | | | | | R1-PM1:V < 132123.6 : 1 (2/0)
| | | | | | | | | | R1-PM1:V >= 132123.6 : 0 (224/0)
| | | | | | R4-PM2:V >= 131987.69


Classifier output
| | | | | R4-PM2:V >= 131987.69
| | | | | | R2-PM10:I < 296.55
| | | | | | | R1-PM1:V < 132512.23 : 1 (138/0)
| | | | | | | R1-PM1:V >= 132512.23
| | | | | | | | R1-PA:ZH < -0.08 : 1 (5/0)
| | | | | | | | R1-PA:ZH >= -0.08 : 0 (98/0)
| | | | | | R2-PM10:I >= 296.55
| | | | | | | R2-PM4:I < 342.66
| | | | | | | | R2-PM3:V < 131365.73 : 1 (189/0)
| | | | | | | | R2-PM3:V >= 131365.73 : 0 (10/0)
| | | | | | | R2-PM4:I >= 342.66 : 1 (250/0)
| | | | R1:DF >= 0.01
| | | | | R4-PM11:I < 1.65
| | | | | | R4:F < 60 : 0 (200/0)
| | | | | | R4:F >= 60
| | | | | | | R4-PM7:V < 132151.31 : 0 (55/0)
| | | | | | | R4-PM7:V >= 132151.31
| | | | | | | | R4:F < 60 : 1 (3/0)
| | | | | | | | R4:F >= 60 : 0 (16/0)
| | | | | R4-PM11:I >= 1.65 : 1 (2/0)
| | | R1-PM3:V >= 133377.26 : 1 (165/0)
| R2-PM11:I >= 3.1
| | R4-PM5:I < 423.44
| | | R2-PA2:VH < 12.96
| | | | R2-PA:ZH < -3.04
| | | | | R4-PA11:IH < 158.88
| | | | | | R4-PM6:I < 258.09 : 1 (145/0)
| | | | | | R4-PM6:I >= 258.09
| | | | | | | R3-PM3:V < 131095.59
| | | | | | | | R2-PA6:IH < -10.56
| | | | | | | | | R1-PM7:V < 132173.74
| | | | | | | | | | R3-PM11:I < 4.12
| | | | | | | | | | | R3-PM12:I < 1.46 : 1 (13/0)
| | | | | | | | | | | R3-PM12:I >= 1.46
| | | | | | | | | | | | R1-PA12:IH < 139.79
| | | | | | | | | | | | | R3-PM11:I < 3.48
| | | | | | | | | | | | | | R1-PM4:I < 316.41 : 1 (2/0)
| | | | | | | | | | | | | | R1-PM4:I >= 316.41


Classifier output
| | | | | | | | | | | | | | | R1-PM4:I >= 316.41
| | | | | | | | | | | | | | | | R3-PA10:IH < 126.74 : 1 (3/0)
| | | | | | | | | | | | | | | | R3-PA10:IH >= 126.74 : 0 (3/0)
| | | | | | | | | | | | | | | R3-PM11:I >= 3.48 : 1 (8/0)
| | | | | | | | | | | | | | R1-PA12:IH >= 139.79 : 0 (7/0)
| | | | | | | | | | | R3-PM11:I >= 4.12
| | | | | | | | | | | | R2-PA12:IH < -27.04
| | | | | | | | | | | | | R3-PA1:VH < -36.46
| | | | | | | | | | | | | | R2-PA11:IH < -177.08
| | | | | | | | | | | | | | | R1-PM11:I < 12.45
| | | | | | | | | | | | | | | | R2-PA1:VH < -48.24 : 1 (11/0)
| | | | | | | | | | | | | | | | R2-PA1:VH >= -48.24 : 0 (1/0)
| | | | | | | | | | | | | | | R1-PM11:I >= 12.45 : 0 (1/0)
| | | | | | | | | | | | | | R2-PA11:IH >= -177.08 : 1 (92/0)
| | | | | | | | | | | | | R3-PA1:VH >= -36.46
| | | | | | | | | | | | | | R1-PM2:V < 131697.35
| | | | | | | | | | | | | | | R4-PA3:VH < 91.72 : 1 (7/0)
| | | | | | | | | | | | | | | R4-PA3:VH >= 91.72
| | | | | | | | | | | | | | | | R1-PA6:IH < 92.67 : 0 (1/0)
| | | | | | | | | | | | | | | | R1-PA6:IH >= 92.67 : 1 (1/0)
| | | | | | | | | | | | | | R1-PM2:V >= 131697.35 : 0 (4/0)
| | | | | | | | | | | | R2-PA12:IH >= -27.04
| | | | | | | | | | | | | R4-PA12:IH < -169.95
| | | | | | | | | | | | | | R2-PM4:I < 368.51 : 1 (15/0)
| | | | | | | | | | | | | | R2-PM4:I >= 368.51 : 0 (1/0)
| | | | | | | | | | | | | R4-PA12:IH >= -169.95 : 1 (220/0)
| | | | | | | | | | | R1-PM7:V >= 132173.74
| | | | | | | | | | | | R3-PM11:I < 7.78
| | | | | | | | | | | | | R2-PA10:IH < -169.48 : 0 (1/0)
| | | | | | | | | | | | | R2-PA10:IH >= -169.48 : 1 (15/0)
| | | | | | | | | | | | R3-PM11:I >= 7.78 : 0 (14/0)
| | | | | | | | | | R2-PA6:IH >= -10.56
| | | | | | | | | | | R2:F < 60
| | | | | | | | | | | | R2-PM11:I < 4.78
| | | | | | | | | | | | | R3-PM2:V < 129666.42 : 1 (6/0)
| | | | | | | | | | | | | R3-PM2:V >= 129666.42
| | | | | | | | | | | | | | R4-PM7:V < 131951.63 : 0 (17/0)
| | | | | | | | | | | | | | R4-PM7:V >= 131951.63

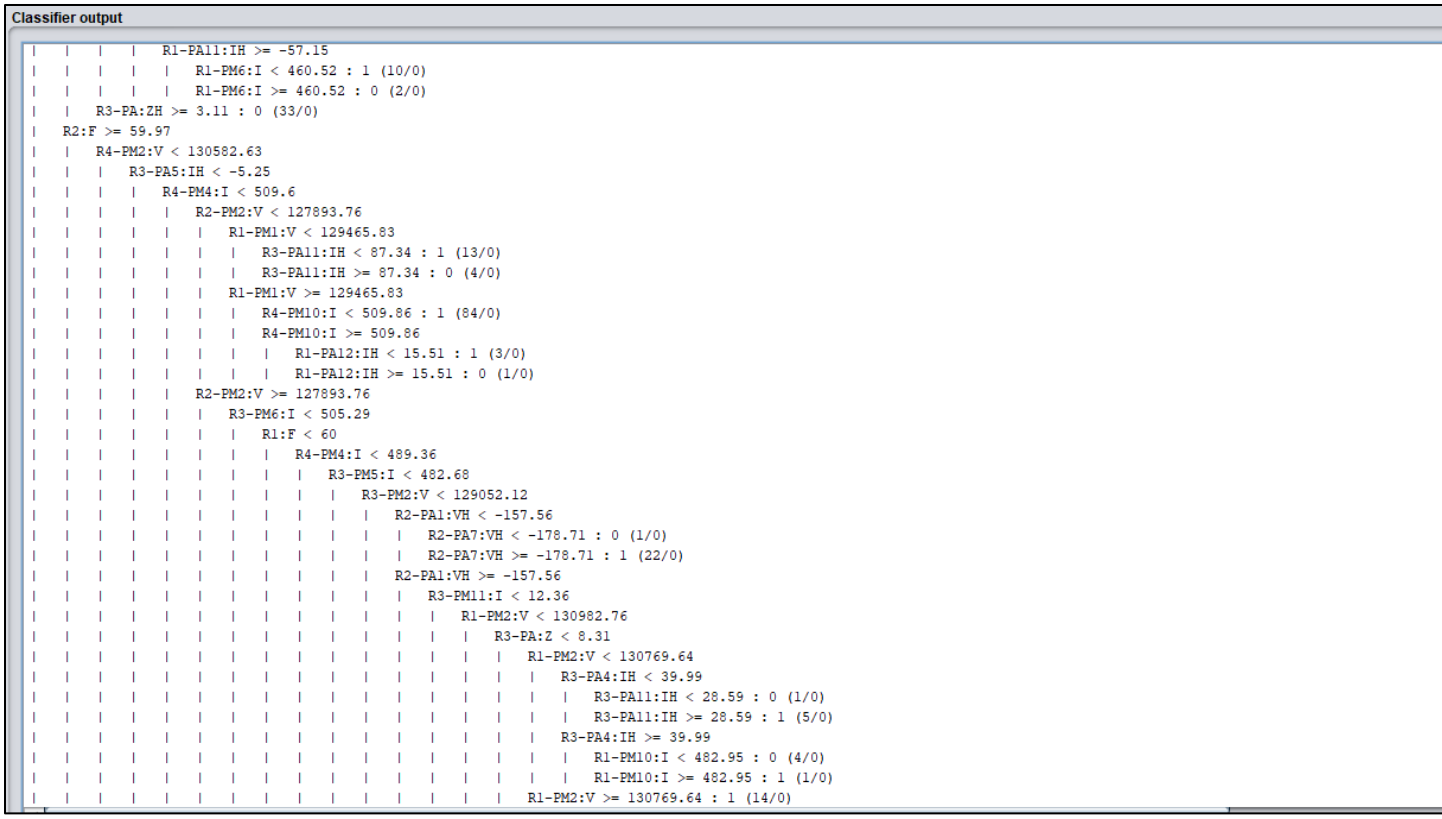

Classifier output
| | | | | | | | | | | | | | R4-PM7:V >= 131951.63
| | | | | | | | | | | | | | | R2-PM7:V < 130987.25 : 1 (7/0)
| | | | | | | | | | | | | | | R2-PM7:V >= 130987.25 : 0 (5/0)
| | | | | | | | | | | | | R2-PM11:I >= 4.78
| | | | | | | | | | | | | | R4-PM2:V < 131845.28
| | | | | | | | | | | | | | R2-PA1:VH < 52.73
| | | | | | | | | | | | | | | R1-PA7:VH < 58.6 : 1 (1/0)
| | | | | | | | | | | | | | | R1-PA7:VH >= 58.6 : 0 (4/0)
| | | | | | | | | | | | | | R2-PA1:VH >= 52.73
| | | | | | | | | | | | | | | R2-PM3:V < 130501.92 : 1 (37/0)
| | | | | | | | | | | | | | | R2-PM3:V >= 130501.92
| | | | | | | | | | | | | | | | R3-PM1:V < 130130.27
| | | | | | | | | | | | | | | | | R2-PM10:I < 370.75 : 1 (1/0)
| | | | | | | | | | | | | | | | | R2-PM10:I >= 370.75 : 0 (6/0)
| | | | | | | | | | | | | | | | R3-PM1:V >= 130130.27
| | | | | | | | | | | | | | | | | R1-PA10:IH < 90.45
| | | | | | | | | | | | | | | | | | R1-PA10:IH < 88.12
| | | | | | | | | | | | | | | | | | | R3-PA:Z < 11.2 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | R3-PA:Z >= 11.2 : 1 (6/0)
| | | | | | | | | | | | | | | | | | R1-PA10:IH >= 88.12 : 0 (3/0)
| | | | | | | | | | | | | | | | | R1-PA10:IH >= 90.45 : 1 (13/0)
| | | | | | | | | | | | | R4-PM2:V >= 131845.28 : 1 (52/0)
| | | | | | | | | | | | R2:F >= 60 : 0 (10/0)
| | | | | | | | | | | R3-PM3:V >= 131095.59
| | | | | | | | | | | | R4-PM4:I < 264.87
| | | | | | | | | | | | | R2-PM12:I < 9.04 : 1 (11/0)
| | | | | | | | | | | | | R2-PM12:I >= 9.04 : 0 (3/0)
| | | | | | | | | | | | R4-PM4:I >= 264.87
| | | | | | | | | | | | | R3-PM3:V < 131998.23
| | | | | | | | | | | | | | R2-PM10:I < 275.5 : 1 (1/0)
| | | | | | | | | | | | | | R2-PM10:I >= 275.5 : 0 (25/0)
| | | | | | | | | | | | | R3-PM3:V >= 131998.23 : 1 (4/0)
| | | | | | | | | | | R4-PA11:IH >= 158.88 : 1 (87/0)
| | | | | | | | | | R2-PA:ZH >= -3.04
| | | | | | | | | | | R1-PA3:VH < -177.94
| | | | | | | | | | | | R1:F < 59.98 : 0 (2/0)
| | | | | | | | | | | | R1:F >= 59.98 : 1 (2/0)
| | | | | | | | | | | R1-PA3:VH >= -177.94

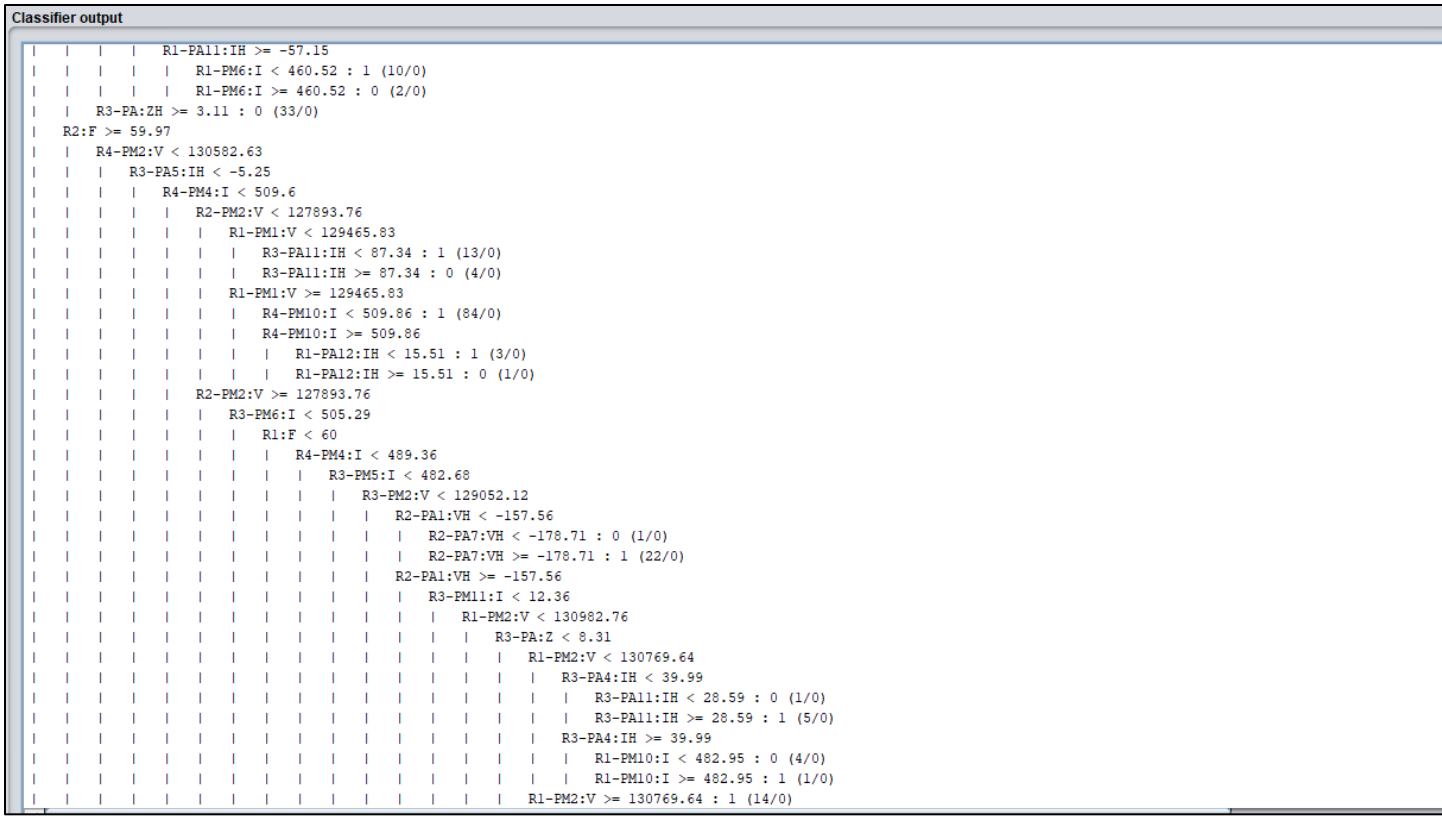

Classifier output
| | | | | R1-PA3:VH >= -177.94
| | | | | | R1-PA3:VH < 177.26 : 1 (367/0)
| | | | | | R1-PA3:VH >= 177.26
| | | | | | | R2-PA5:IH < 107.25
| | | | | | | | R1-PM3:V < 131571.98 : 0 (1/0)
| | | | | | | | R1-PM3:V >= 131571.98 : 1 (2/0)
| | | | | | | R2-PA5:IH >= 107.25 : 1 (8/0)
| | | R2-PA2:VH >= 12.96 : 1 (313/0)
| | R4-PM5:I >= 423.44
| | | R2-PM10:I < 594.34
| | | | R3-PA3:VH < -169.86
| | | | | R1-PM7:V < 131120.67
| | | | | | R3-PA3:VH < -174.84 : 1 (17/0)
| | | | | | R3-PA3:VH >= -174.84
| | | | | | | R1-PM10:I < 462.99 : 1 (5/0)
| | | | | | | R1-PM10:I >= 462.99 : 0 (13/0)
| | | | | R1-PM7:V >= 131120.67 : 0 (22/0)
| | | | R3-PA3:VH >= -169.86
| | | | | R4-PA:ZH < 0.07
| | | | | | R4-PA7:VH < 119.93
| | | | | | | R2-PM3:V < 129653.03
| | | | | | | | R3:F < 60
| | | | | | | | | R1-PA3:VH < 137.4
| | | | | | | | | | R3-PM5:I < 474.99
| | | | | | | | | | | R4-PM7:V < 130230.56
| | | | | | | | | | | | R3-PA10:IH < 141.85
| | | | | | | | | | | | | R3-PM7:V < 127948.9 : 1 (1/0)
| | | | | | | | | | | | | R3-PM7:V >= 127948.9 : 0 (4/0)
| | | | | | | | | | | | R3-PA10:IH >= 141.85 : 1 (10/0)
| | | | | | | | | | | R4-PM7:V >= 130230.56
| | | | | | | | | | | | R4-PA12:IH < -147.4
| | | | | | | | | | | | | R2-PA6:IH < -63.65 : 1 (8/0)
| | | | | | | | | | | | | R2-PA6:IH >= -63.65 : 0 (1/0)
| | | | | | | | | | | | R4-PA12:IH >= -147.4 : 1 (103/0)
| | | | | | | | | | R3-PM5:I >= 474.99
| | | | | | | | | | | R1-PA12:IH < -168.1
| | | | | | | | | | | | R2-PM7:V < 128332.26
| | | | | | | | | | | | | R2-PM4:I < 504.85


Classifier output
| | | | | | | | | | | | | | R2-PM4:I < 504.85
| | | | | | | | | | | | | | | R1-PA2:VH < -127.12 : 1 (1/0)
| | | | | | | | | | | | | | | R1-PA2:VH >= -127.12 : 0 (3/0)
| | | | | | | | | | | | | | R2-PM4:I >= 504.85 : 1 (5/0)
| | | | | | | | | | | | | R2-PM7:V >= 128332.26 : 0 (9/0)
| | | | | | | | | | | | R1-PA12:IH >= -168.1
| | | | | | | | | | | | | R1-PA1:VH < -32.51 : 1 (35/0)
| | | | | | | | | | | | | R1-PA1:VH >= -32.51
| | | | | | | | | | | | | | R3-PA6:IH < -96.91 : 0 (7/0)
| | | | | | | | | | | | | | R3-PA6:IH >= -96.91
| | | | | | | | | | | | | | | R4-PM10:I < 465.7
| | | | | | | | | | | | | | | | R2-PM3:V < 128056.77 : 1 (1/0)
| | | | | | | | | | | | | | | | R2-PM3:V >= 128056.77
| | | | | | | | | | | | | | | | | R1-PM5:I < 474.16 : 1 (1/0)
| | | | | | | | | | | | | | | | | R1-PM5:I >= 474.16 : 0 (4/0)
| | | | | | | | | | | | | | | R4-PM10:I >= 465.7
| | | | | | | | | | | | | | | | R1-PM3:V < 129114.8 : 0 (1/0)
| | | | | | | | | | | | | | | | R1-PM3:V >= 129114.8
| | | | | | | | | | | | | | | | | R1-PM11:I < 3.75 : 0 (1/0)
| | | | | | | | | | | | | | | | | R1-PM11:I >= 3.75
| | | | | | | | | | | | | | | | | | R1-PM6:I < 488.26
| | | | | | | | | | | | | | | | | | | R1-PM7:V < 130995.3
| | | | | | | | | | | | | | | | | | | | R4-PA12:IH < 176.41
| | | | | | | | | | | | | | | | | | | | | R1-PA3:VH < -167.76 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | | | R1-PA3:VH >= -167.76 : 1 (36/0)
| | | | | | | | | | | | | | | | | | | | R4-PA12:IH >= 176.41 : 0 (1/0)
| | | | | | | | | | | | | | | | | | | R1-PM7:V >= 130995.3
| | | | | | | | | | | | | | | | | | | | R1-PM3:V < 131496.76 : 0 (2/0)
| | | | | | | | | | | | | | | | | | | | R1-PM3:V >= 131496.76 : 1 (2/0)
| | | | | | | | | | | | | | | | | | R1-PM6:I >= 488.26 : 1 (50/0)
| | | | | | | | | | R1-PA3:VH >= 137.4 : 1 (70/0)
| | | | | | | | | R3:F >= 60
| | | | | | | | | | R2-PM8:V < 1241.13 : 0 (6/0)
| | | | | | | | | | R2-PM8:V >= 1241.13 : 1 (2/0)
| | | | | | | | R2-PM3:V >= 129653.03
| | | | | | | | | R1-PM10:I < 416.3 : 1 (27/0)
| | | | | | | | | R1-PM10:I >= 416.3
| | | | | | | | | | R4-PA:ZH < -0.03

Classifier output

```
|   |   |   |   |   |   |   |   |   R4-PA:ZH < -0.03
|   |   |   |   |   |   |   |   |   |   R3-PA5:IH < 67.93
|   |   |   |   |   |   |   |   |   |   |   R4-PA:Z < 9.53 : 1 (4/0)
|   |   |   |   |   |   |   |   |   |   |   R4-PA:Z >= 9.53
|   |   |   |   |   |   |   |   |   |   |   |   R2-PA:Z < 9.69 : 0 (5/0)
|   |   |   |   |   |   |   |   |   |   |   |   R2-PA:Z >= 9.69
|   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM12:I < 20.08 : 1 (2/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM12:I >= 20.08 : 0 (2/0)
|   |   |   |   |   |   |   |   |   |   R3-PA5:IH >= 67.93 : 1 (17/0)
|   |   |   |   |   |   |   |   |   R4-PA:ZH >= -0.03
|   |   |   |   |   |   |   |   |   |   R1-PM5:I < 433.7
|   |   |   |   |   |   |   |   |   |   |   R1-PM1:V < 131371.4
|   |   |   |   |   |   |   |   |   |   |   |   R1-PM2:V < 131158.28 : 0 (1/0)
|   |   |   |   |   |   |   |   |   |   |   |   R1-PM2:V >= 131158.28 : 1 (5/0)
|   |   |   |   |   |   |   |   |   |   |   R1-PM1:V >= 131371.4 : 0 (5/0)
|   |   |   |   |   |   |   |   |   |   R1-PM5:I >= 433.7 : 0 (18/0)
|   |   |   |   |   |   R4-PA7:VH >= 119.93
|   |   |   |   |   |   |   R1-PM7:V < 130945.15
|   |   |   |   |   |   |   |   R2-PM11:I < 11.67 : 1 (25/0)
|   |   |   |   |   |   |   |   R2-PM11:I >= 11.67
|   |   |   |   |   |   |   |   |   R4:DF < 0.01 : 0 (8/0)
|   |   |   |   |   |   |   |   |   R4:DF >= 0.01 : 1 (6/0)
|   |   |   |   |   |   |   R1-PM7:V >= 130945.15
|   |   |   |   |   |   |   |   R1-PM6:I < 456.77
|   |   |   |   |   |   |   |   |   R2-PM1:V < 129734.96 : 1 (8/0)
|   |   |   |   |   |   |   |   |   R2-PM1:V >= 129734.96
|   |   |   |   |   |   |   |   |   |   R3-PM1:V < 130055.05 : 0 (10/0)
|   |   |   |   |   |   |   |   |   |   R3-PM1:V >= 130055.05 : 1 (3/0)
|   |   |   |   |   |   |   |   R1-PM6:I >= 456.77 : 0 (17/0)
|   |   |   |   |   R4-PA:ZH >= 0.07 : 1 (95/0)
|   |   |   R2-PM10:I >= 594.34 : 0 (21/0)
R2-PA11:IH >= 1.94
|   R2:F < 59.97
|   |   R3-PA:ZH < 3.11
|   |   |   R3-PA4:IH < 42.32 : 1 (31/0)
|   |   |   R3-PA4:IH >= 42.32
|   |   |   |   R1-PA11:IH < -57.15 : 0 (2/0)
|   |   |   |   R1-PA11:IH >= -57.15
```


Classifier output
| | | | R1-PA11:IH >= -57.15
| | | | | R1-PM6:I < 460.52 : 1 (10/0)
| | | | | R1-PM6:I >= 460.52 : 0 (2/0)
| | R3-PA:ZH >= 3.11 : 0 (33/0)
| R2:F >= 59.97
| | R4-PM2:V < 130582.63
| | | R3-PA5:IH < -5.25
| | | | R4-PM4:I < 509.6
| | | | | R2-PM2:V < 127893.76
| | | | | | R1-PM1:V < 129465.83
| | | | | | | R3-PA11:IH < 87.34 : 1 (13/0)
| | | | | | | R3-PA11:IH >= 87.34 : 0 (4/0)
| | | | | | R1-PM1:V >= 129465.83
| | | | | | | R4-PM10:I < 509.86 : 1 (84/0)
| | | | | | | R4-PM10:I >= 509.86
| | | | | | | | R1-PA12:IH < 15.51 : 1 (3/0)
| | | | | | | | R1-PA12:IH >= 15.51 : 0 (1/0)
| | | | | R2-PM2:V >= 127893.76
| | | | | | R3-PM6:I < 505.29
| | | | | | | R1:F < 60
| | | | | | | | R4-PM4:I < 489.36
| | | | | | | | | R3-PM5:I < 482.68
| | | | | | | | | | R3-PM2:V < 129052.12
| | | | | | | | | | | R2-PA1:VH < -157.56
| | | | | | | | | | | | R2-PA7:VH < -178.71 : 0 (1/0)
| | | | | | | | | | | | R2-PA7:VH >= -178.71 : 1 (22/0)
| | | | | | | | | | | R2-PA1:VH >= -157.56
| | | | | | | | | | | | R3-PM11:I < 12.36
| | | | | | | | | | | | | R1-PM2:V < 130982.76
| | | | | | | | | | | | | | R3-PA:Z < 8.31
| | | | | | | | | | | | | | | R1-PM2:V < 130769.64
| | | | | | | | | | | | | | | | R3-PA4:IH < 39.99
| | | | | | | | | | | | | | | | | R3-PA11:IH < 28.59 : 0 (1/0)
| | | | | | | | | | | | | | | | | R3-PA11:IH >= 28.59 : 1 (5/0)
| | | | | | | | | | | | | | | | R3-PA4:IH >= 39.99
| | | | | | | | | | | | | | | | | R1-PM10:I < 482.95 : 0 (4/0)
| | | | | | | | | | | | | | | | | R1-PM10:I >= 482.95 : 1 (1/0)
| | | | | | | | | | | | | | | R1-PM2:V >= 130769.64 : 1 (14/0)


Classifier output
| | | | R1-PA11:IH >= -57.15
| | | | | R1-PM6:I < 460.52 : 1 (10/0)
| | | | | R1-PM6:I >= 460.52 : 0 (2/0)
| | R3-PA:ZH >= 3.11 : 0 (33/0)
| R2:F >= 59.97
| | R4-PM2:V < 130582.63
| | | R3-PA5:IH < -5.25
| | | | R4-PM4:I < 509.6
| | | | | R2-PM2:V < 127893.76
| | | | | | R1-PM1:V < 129465.83
| | | | | | | R3-PA11:IH < 87.34 : 1 (13/0)
| | | | | | | R3-PA11:IH >= 87.34 : 0 (4/0)
| | | | | | R1-PM1:V >= 129465.83
| | | | | | | R4-PM10:I < 509.86 : 1 (84/0)
| | | | | | | R4-PM10:I >= 509.86
| | | | | | | | R1-PA12:IH < 15.51 : 1 (3/0)
| | | | | | | | R1-PA12:IH >= 15.51 : 0 (1/0)
| | | | | R2-PM2:V >= 127893.76
| | | | | | R3-PM6:I < 505.29
| | | | | | | R1:F < 60
| | | | | | | | R4-PM4:I < 489.36
| | | | | | | | | R3-PM5:I < 482.68
| | | | | | | | | | R3-PM2:V < 129052.12
| | | | | | | | | | | R2-PA1:VH < -157.56
| | | | | | | | | | | | R2-PA7:VH < -178.71 : 0 (1/0)
| | | | | | | | | | | | R2-PA7:VH >= -178.71 : 1 (22/0)
| | | | | | | | | | | R2-PA1:VH >= -157.56
| | | | | | | | | | | | R3-PM11:I < 12.36
| | | | | | | | | | | | | R1-PM2:V < 130982.76
| | | | | | | | | | | | | | R3-PA:Z < 8.31
| | | | | | | | | | | | | | | R1-PM2:V < 130769.64
| | | | | | | | | | | | | | | | R3-PA4:IH < 39.99
| | | | | | | | | | | | | | | | | R3-PA11:IH < 28.59 : 0 (1/0)
| | | | | | | | | | | | | | | | | R3-PA11:IH >= 28.59 : 1 (5/0)
| | | | | | | | | | | | | | | | R3-PA4:IH >= 39.99
| | | | | | | | | | | | | | | | | R1-PM10:I < 482.95 : 0 (4/0)
| | | | | | | | | | | | | | | | | R1-PM10:I >= 482.95 : 1 (1/0)
| | | | | | | | | | | | | | | R1-PM2:V >= 130769.64 : 1 (14/0)

Classifier output

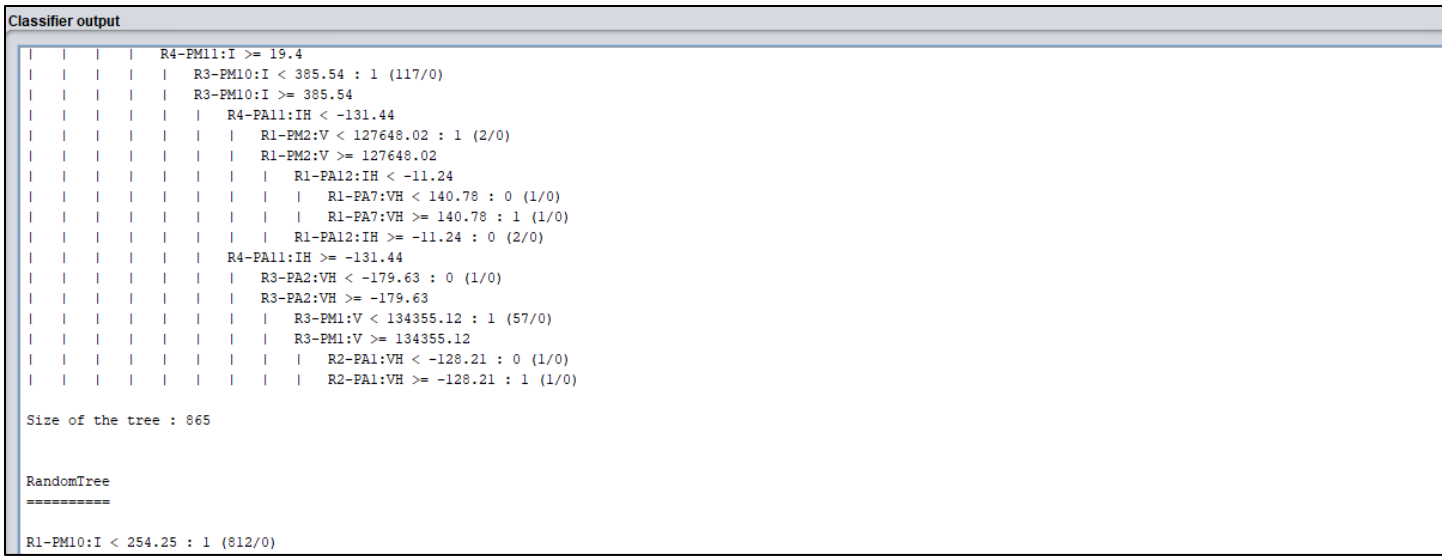
```
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM2:V >= 130769.64 : 1 (14/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PA:Z >= 8.31 : 1 (41/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM2:V >= 130982.76
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM5:I < 456.13 : 1 (4/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM5:I >= 456.13
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM1:V < 129116.78 : 1 (2/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM1:V >= 129116.78 : 0 (19/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM11:I >= 12.36
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PA6:IH < -139.44 : 1 (7/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PA6:IH >= -139.44
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R4-PA12:IH < 46.65
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PA6:IH < 92.89 : 0 (7/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PA6:IH >= 92.89 : 1 (5/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R4-PA12:IH >= 46.65 : 0 (17/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM2:V >= 129052.12 : 1 (32/0)
|   |   |   |   |   |   |   |   |   |   |   |   R3-PM5:I >= 482.68 : 1 (35/0)
|   |   |   |   |   |   |   |   |   |   |   R4-PM4:I >= 489.36
|   |   |   |   |   |   |   |   |   |   |   |   R2-PM10:I < 502.02 : 0 (10/0)
|   |   |   |   |   |   |   |   |   |   |   |   R2-PM10:I >= 502.02
|   |   |   |   |   |   |   |   |   |   |   |   |   R3-PA6:IH < 143.49
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM11:I < 14.01 : 1 (16/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM11:I >= 14.01 : 0 (1/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R3-PA6:IH >= 143.49
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA6:IH < -22.09 : 0 (15/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA6:IH >= -22.09 : 1 (4/0)
|   |   |   |   |   |   |   |   |   |   |   R1:F >= 60 : 0 (15/0)
|   |   |   |   |   |   |   |   |   |   R3-PM6:I >= 505.29 : 1 (24/0)
|   |   |   |   |   R4-PM4:I >= 509.6
|   |   |   |   |   |   R2-PM1:V < 122555.39 : 1 (8/0)
|   |   |   |   |   |   R2-PM1:V >= 122555.39
|   |   |   |   |   |   |   R4-PA3:VH < -63.09
|   |   |   |   |   |   |   |   R1-PM7:V < 128939.29 : 0 (1/0)
|   |   |   |   |   |   |   |   R1-PM7:V >= 128939.29 : 1 (3/0)
|   |   |   |   |   |   |   R4-PA3:VH >= -63.09
|   |   |   |   |   |   |   |   R1-PM12:I < 27.19 : 0 (26/0)
|   |   |   |   |   |   |   |   R1-PM12:I >= 27.19 : 1 (1/0)
|   |   |   R3-PA5:IH >= -5.25 : 1 (76/0)
|   |   R4-PM2:V >= 130582.63
```

Classifier output

```
|   |   R4-PM2:V >= 130582.63
|   |   |   R2-PA:ZH < -3.05
|   |   |   |   R3-PA2:VH < 173.67
|   |   |   |   |   R4-PM10:I < 333.36
|   |   |   |   |   |   R1-PA3:VH < -49.63
|   |   |   |   |   |   |   R3-PA2:VH < 65.66
|   |   |   |   |   |   |   |   R3:F < 60.02
|   |   |   |   |   |   |   |   |   R2-PA11:IH < 6.02
|   |   |   |   |   |   |   |   |   |   R3-PA7:VH < 132.75 : 0 (3/0)
|   |   |   |   |   |   |   |   |   |   R3-PA7:VH >= 132.75 : 1 (3/0)
|   |   |   |   |   |   |   |   |   R2-PA11:IH >= 6.02 : 1 (141/0)
|   |   |   |   |   |   |   |   R3:F >= 60.02 : 0 (1/0)
|   |   |   |   |   |   |   R3-PA2:VH >= 65.66 : 0 (6/0)
|   |   |   |   |   |   R1-PA3:VH >= -49.63 : 1 (431/0)
|   |   |   |   |   R4-PM10:I >= 333.36
|   |   |   |   |   |   R1-PM1:V < 131521.84
|   |   |   |   |   |   |   R2-PM10:I < 431.33 : 1 (233/0)
|   |   |   |   |   |   |   R2-PM10:I >= 431.33
|   |   |   |   |   |   |   |   R2-PM3:V < 129497.7 : 1 (168/0)
|   |   |   |   |   |   |   |   R2-PM3:V >= 129497.7
|   |   |   |   |   |   |   |   |   R2-PM5:I < 464.31
|   |   |   |   |   |   |   |   |   |   R1-PM3:V < 131421.54 : 1 (14/0)
|   |   |   |   |   |   |   |   |   |   R1-PM3:V >= 131421.54 : 0 (2/0)
|   |   |   |   |   |   |   |   |   R2-PM5:I >= 464.31 : 0 (7/0)
|   |   |   |   |   |   R1-PM1:V >= 131521.84
|   |   |   |   |   |   |   R4-PA6:IH < -38.22 : 1 (65/0)
|   |   |   |   |   |   |   R4-PA6:IH >= -38.22
|   |   |   |   |   |   |   |   R1-PM12:I < 4.3
|   |   |   |   |   |   |   |   |   R3-PA:Z < 9.89 : 0 (8/0)
|   |   |   |   |   |   |   |   |   R3-PA:Z >= 9.89
|   |   |   |   |   |   |   |   |   |   R4-PM12:I < 3.75 : 1 (3/0)
|   |   |   |   |   |   |   |   |   |   R4-PM12:I >= 3.75 : 0 (1/0)
|   |   |   |   |   |   |   |   R1-PM12:I >= 4.3
|   |   |   |   |   |   |   |   |   R3-PA2:VH < 135.33
|   |   |   |   |   |   |   |   |   |   R3:F < 60
|   |   |   |   |   |   |   |   |   |   |   R1-PA2:VH < 140.64
|   |   |   |   |   |   |   |   |   |   |   |   R3-PA:Z < 10.46
|   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA3:VH < 16.1
```

Classifier output

```
|   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA3:VH < 16.1
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM2:V < 129917.15
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM5:I < 406.14
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA2:VH < 84.38
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PA3:VH < -42.58 : 1 (1/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R2-PA3:VH >= -42.58 : 0 (1/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA2:VH >= 84.38 : 1 (17/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM5:I >= 406.14 : 0 (3/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   R3-PM2:V >= 129917.15
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM1:V < 133402.33 : 0 (9/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   R1-PM1:V >= 133402.33 : 1 (3/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R1-PA3:VH >= 16.1 : 1 (29/0)
|   |   |   |   |   |   |   |   |   |   |   |   R3-PA:Z >= 10.46
|   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM1:V < 130898.35 : 1 (54/0)
|   |   |   |   |   |   |   |   |   |   |   |   |   R2-PM1:V >= 130898.35 : 0 (2/0)
|   |   |   |   |   |   |   |   |   |   |   |   R1-PA2:VH >= 140.64 : 0 (4/0)
|   |   |   |   |   |   |   |   |   |   |   R3:F >= 60 : 0 (9/0)
|   |   |   |   |   |   |   |   |   |   R3-PA2:VH >= 135.33 : 1 (31/0)
|   |   |   |   |   R3-PA2:VH >= 173.67 : 0 (7/0)
|   |   |   R2-PA:ZH >= -3.05
|   |   |   |   R4-PM11:I < 19.4 : 1 (586/0)
|   |   |   |   R4-PM11:I >= 19.4
```



Classifier output

```
|   |   |   |   R4-PM11:I >= 19.4
|   |   |   |   |   R3-PM10:I < 385.54 : 1 (117/0)
|   |   |   |   |   R3-PM10:I >= 385.54
|   |   |   |   |   |   R4-PA11:IH < -131.44
|   |   |   |   |   |   |   R1-PM2:V < 127648.02 : 1 (2/0)
|   |   |   |   |   |   |   R1-PM2:V >= 127648.02
|   |   |   |   |   |   |   |   R1-PA12:IH < -11.24
|   |   |   |   |   |   |   |   |   R1-PA7:VH < 140.78 : 0 (1/0)
|   |   |   |   |   |   |   |   |   R1-PA7:VH >= 140.78 : 1 (1/0)
|   |   |   |   |   |   |   |   R1-PA12:IH >= -11.24 : 0 (2/0)
|   |   |   |   |   |   R4-PA11:IH >= -131.44
|   |   |   |   |   |   |   R3-PA2:VH < -179.63 : 0 (1/0)
|   |   |   |   |   |   |   R3-PA2:VH >= -179.63
|   |   |   |   |   |   |   |   R3-PM1:V < 134355.12 : 1 (57/0)
|   |   |   |   |   |   |   |   R3-PM1:V >= 134355.12
|   |   |   |   |   |   |   |   |   R2-PA1:VH < -128.21 : 0 (1/0)
|   |   |   |   |   |   |   |   |   R2-PA1:VH >= -128.21 : 1 (1/0)

Size of the tree : 865


RandomTree
==========

R1-PM10:I < 254.25 : 1 (812/0)
```

```
Attribute importance based on average impurity decrease (and number of nodes using that attribute)

      0.55 (     1)  relay1_log
      0.54 (     2)  R2-PM8:V
      0.5  (     3)  R1-PM8:V
      0.45 (     1)  R4-PM8:V
      0.45 (   947)  R1-PA1:VH
      0.42 (   813)  R1-PA4:IH
      0.41 (   757)  R1-PM3:V
      0.41 (   964)  R1-PA2:VH
      0.4  (   922)  R1-PM4:I
      0.4  (   623)  R1-PA10:IH
      0.4  (   809)  R1-PM1:V
      0.4  (   743)  R1-PA5:IH
      0.39 (   896)  R1-PM5:I
      0.39 (   961)  R1-PA3:VH
      0.39 (   803)  R1-PM6:I
      0.39 (   980)  R1-PM2:V
      0.39 (   744)  R1-PA6:IH
      0.38 (   711)  R1-PA7:VH
      0.38 (   223)  R1-PA11:IH
      0.38 (   605)  R2-PA1:VH
      0.37 (   281)  R1-PM11:I
      0.37 (   787)  R1-PM10:I
      0.36 (   626)  R2-PA2:VH
      0.35 (   268)  R1-PA12:IH
      0.35 (   678)  R1-PA:Z
      0.35 (   681)  R1-PM7:V
      0.34 (   480)  R2-PA7:VH
      0.33 (   580)  R2-PA3:VH
      0.33 (   547)  R2-PM6:I
      0.33 (   178)  R2-PA12:IH
      0.33 (   755)  R2-PM1:V
      0.33 (   606)  R2-PM4:I
      0.32 (   719)  R2-PM2:V
      0.32 (   538)  R2-PA6:IH
      0.32 (   658)  R2-PM5:I
```

```
0.31 (   470)  R3-PA1:VH
0.31 (   800)  R1-PA:ZH
0.31 (   277)  R1-PM12:I
0.3  (   580)  R2-PA4:IH
0.3  (   572)  R2-PA5:IH
0.3  (   770)  R2-PM3:V
0.29 (   524)  R2-PA:Z
0.29 (   633)  R2-PM7:V
0.29 (   520)  R2-PM10:I
0.29 (     5)  R1:S
0.29 (   383)  R3-PM2:V
0.29 (   384)  R1:F
0.29 (   110)  R1:DF
0.28 (   385)  R3-PM1:V
0.28 (   476)  R2-PA10:IH
0.28 (   356)  R3-PM3:V
0.27 (   203)  R2-PM12:I
0.27 (   413)  R3-PA3:VH
0.27 (   875)  R2-PA:ZH
0.27 (   227)  R2-PM11:I
0.27 (   426)  R3-PM6:I
0.27 (   468)  R3-PM4:I
0.27 (   485)  R3-PM5:I
0.26 (   423)  R3-PA2:VH
0.26 (   397)  R3-PA6:IH
0.26 (   322)  R3-PA7:VH
0.26 (   405)  R3-PA4:IH
0.26 (   462)  R3-PA:Z
0.26 (   126)  R3-PA12:IH
0.25 (   431)  R3-PM10:I
0.25 (   277)  R3-PM7:V
0.25 (   402)  R3-PA5:IH
0.25 (   116)  R3-PA11:IH
0.24 (   382)  R4-PM4:I
0.24 (   364)  R4-PA:Z
0.24 (   344)  R4-PM6:I
0.24 (   750)  R3-PA:ZH
```

```
    0.24 (   180)  R2-PA11:IH
    0.24 (   273)  R4-PA4:IH
    0.24 (   283)  R4-PA1:VH
    0.23 (   372)  R3-PA10:IH
    0.23 (    67)  R3:DF
    0.23 (   319)  R2:F
    0.23 (   458)  R4-PM5:I
    0.23 (   344)  R4-PA5:IH
    0.22 (   679)  R4-PM2:V
    0.22 (   299)  R4-PA6:IH
    0.22 (   124)  R4-PA12:IH
    0.22 (   485)  R4-PM3:V
    0.22 (   668)  R4-PA:ZH
    0.22 (   471)  R4-PM1:V
    0.22 (     1)  R2-PM9:V
    0.22 (     4)  R2:S
    0.22 (   229)  R4-PA7:VH
    0.21 (    80)  R2:DF
    0.21 (   337)  R4-PA2:VH
    0.21 (   163)  R3-PM11:I
    0.21 (   370)  R4-PA3:VH
    0.21 (   179)  R3-PM12:I
    0.21 (   523)  R4-PM7:V
    0.2  (   322)  R4-PM10:I
    0.2  (   224)  R4-PA10:IH
    0.2  (     1)  relay2_log
    0.2  (   144)  R4-PM11:I
    0.19 (   294)  R3:F
    0.19 (   289)  R4:F
    0.19 (   132)  R4-PM12:I
    0.18 (   118)  R4-PA11:IH
    0.15 (    74)  R4:DF
    0.14 (     1)  R3-PM8:V
    0.13 (     3)  R3:S
    0.11 (     7)  R4:S
    0.11 (     1)  R2-PA8:VH
    0.11 (     1)  R1-PA8:VH
```

```
0     (     0)  snort_log3
0     (     0)  control_panel_log3
0     (     0)  control_panel_log4
0     (     0)  relay4_log
0     (     0)  snort_log1
0     (     0)  control_panel_log2
0     (     0)  snort_log2
0     (     0)  relay3_log
0     (     0)  R2-PA9:VH
0     (     0)  R3-PA8:VH
0     (     0)  control_panel_log1
0     (     0)  R4-PM9:V
0     (     0)  R4-PA9:VH
0     (     0)  snort_log4
0     (     0)  R3-PA9:VH
0     (     0)  R3-PM9:V
0     (     0)  R1-PM9:V
0     (     0)  R1-PA9:VH
0     (     0)  R4-PA8:VH
```

```
=== Evaluation on test set ===

Time taken to test model on supplied test set: 0.08 seconds

=== Summary ===

Correctly Classified Instances        1293               98.627  %
Incorrectly Classified Instances        18                1.373  %
Kappa statistic                          0.9692
Mean absolute error                      0.056
Root mean squared error                  0.1221
Relative absolute error                 12.5352 %
Root relative squared error             25.7954 %
Total Number of Instances             1311

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.968    0.005    0.991      0.968    0.979      0.969    0.998     0.997     0
                 0.995    0.032    0.984      0.995    0.990      0.969    0.998     0.999     1
Weighted Avg.    0.986    0.022    0.986      0.986    0.986      0.969    0.998     0.998

=== Confusion Matrix ===

   a   b   <-- classified as
 430  14 |   a = 0
   4 863 |   b = 1
```

REFERENCES

Ablebits Software. (March, 2015). Ablebits Ultimate Suite for Microsoft Excel -14 Day Free Trial. *Ablebits*. Retrieved from https://www.ablebits.com/office-addins-blog/2018/01/31/excel-random-selection-random-sample/.

Adhikari, U., Pan, S., Morris, T., Borges-Hink, R., and Beaver, J. (2014). Dataset 1: Power System Datasets. *Misissipi State University in collaboration with Oak Ridge National Laboratories*. Retrieved from https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets.

Adhikari, U., Pan, S., Morris, T., Borges-Hink, R., and Beaver, J. (2014). PowerSystem_Dataset_README. *Misissipi State University in collaboration with Oak Ridge National Laboratories*. Retrieved from https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium on Information Theory.* (pp. 267-281). Akademiai Kiado.

Anderson, T. (2018). Multiple telephonic and email conversations regarding GINA application to thesis from 1AUG17-20JUL18.

Author Unknown (United States Computer Emergency Response Readiness Team). (2018, March). Alert (TA18-074A) – Russian Government Cyber Activity Targeting Energy and Other Critical Infrastructure Sectors. *US-CERT Alerts.* Retrieved from https://www.us-cert.gov/ncas/alerts/TA18-074A.

Author Unknown (User Name: JasonAizkalns). (2014). Identifying specific differences between two datasets in R. *Stack Overflow.* Retrieved from https://stackoverflow.com/questions/27429582/identifying-specific-differences-between-two-data-sets-in-r.

Author Unknown (User Name: Mollie). (2013, September). Using colClasses to load data more quickly in R. *R-Bloggers*. . Retrieved from https://www.r-bloggers.com/using-colclasses-to-load-data-more-quickly-in-r/.

Author Unknown. (Date Unknown). Import, export, and convert data files. *Comprehensive R Archive Network.* Retrieved from https://cran.r-project.org/web/packages/rio/vignettes/rio.html.

Author Unknown. (December, 2012). Understanding the Grid. *North American Electric Reliability Corporation*. Retrieved from http://www.nerc.com/AboutNERC/Documents/Understanding%20the%20Grid%20DEC12.pdf.

Author Unknown. (January, 2014). Thread - Determine the data types of a data frame's columns. *Stack Overflow*. Retrieved from https://stackoverflow.com/questions/21125222/determine-the-data-types-of-a-data-frames-columns%20##%20https://stackoverflow.com/questions/28895044/list-output-truncated-how-to-expand-listed-variables-with-str-in-r.

Author Unknown. (March, 2015). Thread – list output tuncated – How to expand listed variables with str() in R. *Stack Overflow*. Retrieved from https://stackoverflow.com/questions/28895044/list-output-

truncated-how-to-expand-listed-variables-with-str-in-r.

Barnett, D., and Bjornsgaard, K. (2000). *Electric power generation: A nontechnical guide*. Tulsa, OK: PennWell.

Boone, E. (2010). Importing data into R. *YouTube Video Tutorial.*. Retrieved from https://www.youtube.com/watch?v=jGPl8I6lSsM.

Borges-Hink, R. Multiple email conversations regarding associated dataset and research content from 9MAY18-11JUN18..

Borges-Hink, R., Beaver, J., Buckner, M., Morris, T., Adhikari, U., and Pan, S., (2013). An Evaluation of Machine Learning Methods to Detect Malicious SCADA Communications. *2013 12th International Conference on Machine Learning and Applications,* 1-6. Retrieved from http://ieeexplore.ieee.org/document/6786081/.

Borges-Hink, R., Beaver, J., Buckner, M., Morris, T., Adhikari, U., and Pan, S., (2014). Machine Learning for Power System Disturbance and Cyber-attack Discrimination. *Proceedings of the 7th International Symposium on Resilient Control Systems,* 1-8. Retrieved from http://ieeexplore.ieee.org/document/6900095/.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, *52*(3), 345-370. Retrieved from https://pdfs.semanticscholar.org/b8d5/dfa12f50424d8e883663570425d29569cac4.pdf.

Branco, P., Torgo, L., Ribeiro, R. P., Frank, E., Pfahringer, B., and Rau, M. M. (2017, October). Learning Through Utility Optimization in Regression Tasks. In *Data Science and Advanced Analytics (DSAA), 2017 IEEE International Conference on* (pp. 30-39). IEEE. Retrieved from https://www.cs.waikato.ac.nz/ml/publications.html.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32. Retrieved from https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf.

Brownlee, J. (2016, July). How to use machine learning algorithms in Weka. *Machine Learning Mastery.* Retrieved from https://machinelearningmastery.com/use-machine-learning-algorithms-weka/

Caulkins, B. Multiple face to face and email conversations regarding general thesis advice, formation, and content from 22FEB17-20JUL18.

Cisco. (2018). 2018 Annual Cybsercurity Report. *Cisco*. Retrieved from https://www.cisco.com/c/dam/m/digital/elq-cmcglobal/witb/acr2018/acr2018final.pdf?dtid=odicdc000016andccid=cc000160andoid=anrsc005679andecid=8196andelqTrackId=686210143d34494fa27ff73da9690a5bandelqaid=9452andelqat=2.

Coghlan, Avril. (2010). Using R For Multivariate Analysis. *Little Book Of R For Multivariate Analysis*. Retrieved from https://little-book-of-r-for-multivariate-analysis.readthedocs.io/en/latest/src/multivariateanalysis.html.

Cohort 19, Team Bravo. (2013, June). Viable Short-Term Directed Energy Weapon Naval Solutions: A System Analysis of Current Prototypes. *Systems Engineering Capstone Report, Naval Postgraduate School*. Retrieved from https://calhoun.nps.edu/handle/10945/34734.

Crappe, M. (2008). Electric power systems. London, UK. ISTE Ltd.

De Jonge, E., and Van der Loo, E., (2013). An introduction to data cleaning with R. *Statistics Netherland,* 1-53. Retrieved from https://cran.r-project.org/doc/contrib/de_Jonge+van_der_Loo-Introduction_to_data_cleaning_with_R.pdf.

Department of Defense. (2017, August). DoD Initiates Elevation Process for U.S. Cyber Command to a Unified Combatant Command. Retrieved from https://www.defense.gov/News/News-Releases/News-Release-View/Article/1282920/dod-initiates-elevation-process-for-us-cyber-command-to-a-unified-combatant-com/.

Dolk, D., Busalacchi, F., Anderson, T., and Tinsley, D. (2012). GINA: System Interoperability for Enabling Smart Mobile System Services in Network Decision Support Systems. *45th Annual Hawaii International Conference on System Sciences, Computer Society Press.* Retrieved from https://pdfs.semanticscholar.org/7729/852b92a7809f30124d8c5dadca6dc8df6dcd.pdf.

Dougherty, K. (2017). GINA: Identificatino of low-latency obfucsated traffic using multi-attribute analysis. *Naval Postgradaute School Thesis.*

Geilke, M., Karwath, A., Frank, E., and Kramer, S. (2017). Online estimation of discrete, continuous, and conditional joint densities using classifier chains. *Data Mining and Knowledge Discovery*, 1-43. Retreived from https://link.springer.com/epdf/10.1007/s10618-017-0546-6?author_access_token=Tq6WNzJlUCgMEuanx5FYJ_e4RwlQNchNByi7wbcMAY7pxz0FZJ0uHPQKVqQ_KmXzIZcNIfIarCYuONXavH-IQL98gpAEJ0y0_C83BQZC0EwzTwhSHLStPpeoSJXsM5LbqYtIpTVkd9X4LnPS1oTphw%3D%3D.

Gurulian, I., Shepherd, C., Frank, E., Markantonakis, K., Akram, R. N., and Mayes, K. (2017, August). On the Effectiveness of Ambient Sensing for Detecting NFC Relay Attacks. In *Trustcom/BigDataSE/ICESS, 2017 IEEE* (pp. 41-49). IEEE. Retrieved from https://www.cs.waikato.ac.nz/~eibe/pubs/PID4857507.pdf.

Grandoni, D.. (2018, March). The Energy 202: Rick Perry has a plan for Russia's grid attacks. Some lawmakers worry it's not enough.. The *Washington Post*. Retrieved from https://www.washingtonpost.com/news/powerpost/paloma/the-energy-202/2018/03/21/the-energy-202-rick-perry-has-a-plan-for-russia-s-grid-attacks-some-lawmakers-worry-it-s-not-enough/5ab17e7a30fb045e48d05aeb/?utm_term=.636f68a99137.

Greenberg, A. (2017, August). How An Entire Nation Became Russia's Test Lab for Cyberwar. *Wired*. Retrieved from https://www.wired.com/story/russian-hackers-attack-ukraine/.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, *11*(1), 10-18. Retrieved from https://www.cs.waikato.ac.nz/~eibe/pubs/weka_update.pdf.

Holmes, S. (2000, November). RMS Error. Retrieved from http://statweb.stanford.edu/~susan/courses/s60/split/node60.html.

Huda, S., Abawajy, J., Abdollahian, M., Islam, R., and Yearwood, J. (2017). A fast malware feature selection approach using a hybrid of multi-linear and stepwise binary logistic regression. *Concurrency and Computation: Practice and Experience*, *29*(23). Retrieved from https://onlinelibrary-wiley-com.ezproxy.net.ucf.edu/doi/epdf/10.1002/cpe.3912.

ICS-CERT. (2016). ICS-CERT Year In Review. *ICS-CERT.* Retrieved from https://ics-cert.us-cert.gov/sites/default/files/Annual_Reports/Year_in_Review_FY2016_Final_S508C.pdf.

Jamei, M., Stewart, E., Peisert, S., Scaglione, A., McParland, C., Roberts, C., and McEachern, A. (2016). Micro synchrophasor-based intrusion detection in automated distribution systems: Toward critical infrastructure security. *IEEE Internet Computing*, *20*(5), 18-27. Retrieved from https://cloudfront.escholarship.org/dist/prd/content/qt84j8f0md/qt84j8f0md.pdf.

Kabacoff, R. (2017). Bar Plots. *Quick-R Website*. . Retrieved from https://www.statmethods.net/graphs/bar.html.

Khan, R., Albalushi, A., McLaughlin, K., Laverty, D., and Sezer, S. (2017). Model based Intrusion Detection System for Synchrophasor Applications in Smart Grid. Retrieved from https://pure.qub.ac.uk/portal/files/120788299/pesGM2017.pdf.

Lathrop, S. Multiple telephonic and email conversations regarding general thesis advice, formation, and content from 5OCT17-27JUN18.

Mahmoud, M. (Date Unknown). Evaluation of Accuracy of Estimation Methods for Replacing Missing Values for Time Series. *Benha University*. . Retrieved from http://www.bu.edu.eg/portal/uploads/openLearning/Evaluation%20of%20Accuracy%20of%20the%20Estimation%20Methods%20for%20Replacing%20Missing%20Values%20for%20Time%20Series%20Variables.%20Using%20the%20statistical%20packages%20software%20%20(SPSS%20andMTNITAB1%20_paper_en.pdf.

Martin, B. (1995). Instance-based learning: nearest neighbor with generalization [thesis]. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=B7A28775298E320996FA2ED29741E916?doi=10.1.1.588.2069andrep=rep1andtype=pdf.

Metz, T., Pounds, J., and Waters, K. (2015, May). Review, evaluation, and discussion of missing value imputation for mass spectrometry-based label-free global proteomics. *Journal of proteome researrch, 14(5),* 1993-2001. Retreived from doi:10.1109/IIH-MSP.2015.110 Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4776766/.

Microsoft Excel Software. (1987 (initial release)). Microsoft Corporation. Can be accessed/downloaded from https://office.microsoft.com/excel.

Modi, M. U., & Jain, A. (2015). A survey of IDS classification using KDD CUP 99 dataset in WEKA. *Int. J. Sci. Eng. Res*, *6*(11), 947-954. Retrieved from https://pdfs.semanticscholar.org/f3f1/c3919d1c44f3272e7547f02348ba8f25d390.pdf.

Mohammad, S. M., and Bravo-Marquez, F. (2017). Emotion intensities in tweets. *arXiv preprint arXiv:1708.03696*. Retrieved from https://arxiv.org/pdf/1708.03696.pdf.

Morris, T. Multiple email conversations regarding dataset formation and interpretation script from 23JAN18-25JUN18.

Murphy, D. (2017, October). Industrial Control System Network Analysis (Powerpoint Presentation) - 1 Day Course at University of Central Florida. *Security Matters*.

National Institute of Standards and Technology (NIST) . (2015, May). NIST Special Publication 800-82 Revision 2 – Guide to Industrial Control Systems (ICS) Security. Retrieved from

https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-82r2.pdf.

National Institute of Standards and Technology (NIST) - The Smart Grid Interoperability Panel – Cyber Security Working Group. (2010, September). Introduction to NISTIR 7628 Guidelines for Smart Grid Cybersecurity. Retrieved from https://www.nist.gov/document-13017.

Nevlud, P., Bures, M., Kapicak, L., & Zdrálek, J. (2013). Anomaly-based network intrusion detection methods. *Advances in Electrical and Electronic Engineering*, *11*(6), 468. Retrieved from http://search.proquest.com/openview/c66638f8a947e44a3c85548c892c70b8/1?pq-origsite=gscholar&cbl=1616344.

Nguyen, H. A., & Choi, D. (2008, October). Application of data mining to network intrusion detection: classifier selection model. In *Asia-Pacific Network Operations and Management Symposium* (pp. 399-408). Springer, Berlin, Heidelberg. Retrieved from https://link.springer.com/chapter/10.1007/978-3-540-88623-5_41.

Online Trust Aliance. (2018, January). Cyber Incident and Breach Trends Report. *Online Trust Alliance*. Retrieved from https://www.otalliance.org/system/files/files/initiative/documents/ota_cyber_incident_trends_report_jan2018.pdf.

Pan, S., Morris, T., and Adhikari, U. (2015, June). Classification of Distubances and Cyber-Attacks in Power Systems Using Heterogenous Time-Synchornized Data. *IEEE Transactions of Industrial Informatics, 11(3)*, 650-662. . Retrieved from http://ieeexplore.ieee.org/document/7081776/.

Pan, S., Morris, T., and Adhikari, U. (2015, March). A Specification-based Intrusion Detection Framework for Cyber-physical Environment in Electric Power System. *International Journal of Network Security, 17*, 174-188. . Retrieved from http://ijns.jalaxy.com.tw/contents/ijns-v17-n2/ijns-2015-v17-n2-p174-188.pdf.

Perlroth, N. and Sanger, D. (2018, March 15). Cyberattacks Put Russian Giners on the Switch at Power Plants, U.S. Says. The New York Times. Retrieved from https://www.nytimes.com/2018/03/15/us/politics/russia-cyberattacks.html.

Pomerleau, Mark. (2017). Here's how DoD organizes its cyber warriors. Fifth Domain. Retrieved from https://www.fifthdomain.com/workforce/career/2017/07/25/heres-how-dod-organizes-its-cyber-warriors/.

Redwood, O. (2016). Cyber Physical System Vulnerabilty Research. Dissertation submitted/accepted at Florida State University.

Redwood, O. Multiple telephonic and email conversations regarding general ICS information, attack analysis, and attack tactics/techniques from 30OCT17-25JUN18.

RStudio Inc. (2011, February (initial relase)). RStudio Software. Retrieved from https://www.rstudio.com/.

Salmon, D., Zeller, M., Guzman, A., Mynam, V., and Donolo, M. (2009, October). Mitigating the aurora vulnerability with existing technology. In *proceedings of the 36th Annual Western Protective Relay Conference*. Retrieved from https://cdn.selinc.com/assets/Literature/Publications/Technical%20Papers/6392_MitigatingAurora_MZ_20090918_Web.pdf.

Shtatland, E. S., Cain, E., and Barton, M. B. (2001, April). The perils of stepwise logistic regression and how to escape them using information criteria and the output delivery system. In *Proceedings from the 26th Annual SAS Users Group International Conference* (pp. 22-25). Retrieved from http://www2.sas.com/proceedings/sugi26/p222-26.pdf.

Swearingen, M., Brunasso, S., Weiss, J., and Huber, D. (2013). What you need to know (and don't) about the Aurora vulnerability. *Power*, *157*(9), 52-52. Retrieved from http://www.powermag.com/what-you-need-to-know-and-dont-about-the-aurora-vulnerability/.

Symantec. (2017, October). Dragonfly: Western energy sector targeted by sophisticated attack group. *Symantec Corporation*. Retrieved from https://www.symantec.com/blogs/threat-intelligence/dragonfly-energy-sector-cyber-attacks.

Tatsis, V. A., Tjortjis, C., and Tzirakis, P. (2013). Evaluating data mining algorithms using molecular dynamics trajectories. *International journal of data mining and bioinformatics*, *8*(2), 169-187. Retrieved from http://www.ihu.edu.gr/tjortjis/Evaluating%20data%20mining%20algorithms%20using%20molecular%20dynamics%20trajectories%20IJDMB%208_2_Paper%203.pdf.

University of Waikato. (1999 (initial release). Waikato Environment for Knowledge Analysis (WEKA) Version 3.8.1 – Machine Learning and Data Mining Software. *University of Waikaito*. Retreived from https://www.cs.waikato.ac.nz/ml/weka/downloading.html.

University of Waikato. (1999-2018). Publication Page. *University of Waikaito Website*. Retreived from https://www.cs.waikato.ac.nz/ml/publications.html.

Webb-Robertson, B., Wilber, H., Matzke, M., Brown, J., Wang, J., McDermott, J., Smith, R., Rodland, K., Metz, T., Pounds, J., and Waters, K. (2015, May). Review, evaluation, and discussion of missing value imputation for mass spectrometry-based label-free global proteomics. *Journal of proteome researrch, 14(5),* 1993-2001. Retreived from doi:10.1109/IIH-MSP.2015.110 Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4776766/.

Waltman, C. (2016). Aurora: homeland security's secret project to change how we think about cybersecurity. *Muckrock*. Retreived from https://www.muckrock.com/news/archives/2016/nov/14/aurora-generator-test-homeland-security/.

Whitten, I. (Date Unknown). Data Mining With Weka - Webcourse. *University of Waikaito*. Retreived from https://www.futurelearn.com/courses/data-mining-with-weka/.

Whitten, I. (Date Unknown). More Data Mining With Weka - Webcourse. *University of Waikaito*. Retreived from https://www.futurelearn.com/courses/more-data-mining-with-weka.

Wiegand, R. Multiple face to face and email conversations regarding general thesis advice, data analysis, data visualization, and R script construction from 5OCT17-27JUN18.

Wold, S., Esbensen, K., Geladi, P. (1987). Cyber-Physical Attack-Oriented Industrial Control Systems (ICS) Modeling, Analysis and Experiment Environment. (2015). *Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems, 2,* 37-52. Retreived from doi:10.1109/IIH-MSP.2015.110 Retrieved from http://files.isec.pt/DOCUMENTOS/SERVICOS/BIBLIO/Documentos%20de%20acesso%20remoto/Principal%20components%20analysis.pdf.

Yang, Y., McLaughlin, K., Sezer, S., Littler, T., Pranggono, B., Brogan, P., and Wang, H. F. (2013). Intrusion detection system for network security in synchrophasor systems. Retrieved from http://www.academia.edu/download/40457927/19_Intrusion_Detection_System_for_Network_Security_in_Synchrophasor_Systems_ietict13.pdf.

Zeller, M. (2011, February). Common questions and answers addressing the aurora vulnerability. *Schweitzer Engineering Laboratories Technical Report*. Retrieved from https://cdn.selinc.com/assets/Literature/Publications/Technical%20Papers/6467_CommonQuestions_MZ_20101209_Web.pdf?v=20150812-081908.

Zeller, M. (2011, April). Myth or reality—Does the Aurora vulnerability pose a risk to my generator? In *Protective Relay Engineers, 2011 64th Annual Conference for* (pp. 130-136). IEEE. Retrieved from https://ieeexplore-ieee-org.ezproxy.net.ucf.edu/stamp/stamp.jsp?tp=andarnumber=6035612.